GenCore version 5.1.6

OM nucleic - nucleic search, using sw model

Run on: February 19, 2004, 21:05:58 ; Search time 1571 Seconds
(without alignments)
324.885 Million cell updates/sec

Title: US-10-085-108-21_COPY_711_731
Perfect score: 21
Sequence: 1 AAGCCTGATTGATGACCAGGG 21

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 22781392 seqs, 12152238056 residues

Total number of hits satisfying chosen parameters: 45562784

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : EST:*
1: em_estba:*
2: em_esthum:*
3: em_estin:*
4: em_estmu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_htc:*
9: gb_est1:*
10: gb_est2:*
11: gb_htc:*
12: gb_est3:*
13: gb_est4:*
14: gb_est5:*
15: em_estfun:*
16: em_estom:*
17: em_gss_hum:*
18: em_gss_inv:*
19: em_gss_pln:*
20: em_gss_vrt:*
21: em_gss_fun:*
22: em_gss_mam:*
23: em_gss_mus:*
24: em_gss_pro:*
25: em_gss_rod:*
26: em_gss_phg:*
27: em_gss_vrl:*
28: gb_gss1:*
29: gb_gss2:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	18.4	87.6	686	29	CC270269	CC270269 CH261-1N2
	2	18	85.7	809	29	BZ721194	BZ721194 PUCFH73TD
c	3	17.8	84.8	307	9	AA894450	AA894450 oe79f07.s
c	4	17.8	84.8	339	9	AI572687	AI572687 tr78b10.x
	5	17.8	84.8	380	28	BH323391	BH323391 CH230-53J
c	6	17.8	84.8	545	14	CA941354	CA941354 ir34b01.x
c	7	17.8	84.8	592	9	AI816423	AI816423 au47g07.x
c	8	17.8	84.8	671	29	AG121419	AG121419 Pan trogl
c	9	17.8	84.8	708	29	AG186976	AG186976 Pan trogl
	10	17.8	84.8	913	13	BU245702	BU245702 603783316
	11	17.8	84.8	963	10	BG119393	BG119393 602349076
c	12	17.8	84.8	1201	9	AL523235	AL523235 AL523235
	13	17.4	82.9	565	28	AZ072191	AZ072191 RPCI-23-3
c	14	17.4	82.9	914	10	BE729033	BE729033 601561212
c	15	17	81.0	543	9	AI818091	AI818091 wk27a09.x
	16	17	81.0	733	14	CB983939	CB983939 AGENCOURT
	17	17	81.0	763	14	CD101012	CD101012 AGENCOURT
	18	16.8	80.0	321	9	AA967543	AA967543 ua07e06.r
	19	16.8	80.0	341	9	AA118769	AA118769 mp61e01.r
c	20	16.8	80.0	348	29	AG236818	AG236818 Lotus jap
	21	16.8	80.0	359	10	BE861602	BE861602 UI-M-AH1-
c	22	16.8	80.0	364	28	BH056156	BH056156 RPCI-24-3
	23	16.8	80.0	369	9	AW229797	AW229797 uo61h04.y
	24	16.8	80.0	409	9	AA636853	AA636853 vr19h04.r
c	25	16.8	80.0	420	10	BE181117	BE181117 CM3-HT062
	26	16.8	80.0	457	9	AI225508	AI225508 ue88d11.y
	27	16.8	80.0	457	10	BF322594	BF322594 maa39b08.
	28	16.8	80.0	497	10	BG276054	BG276054 uu89e08.y
	29	16.8	80.0	509	9	AW230744	AW230744 uo67d02.y
	30	16.8	80.0	512	13	BU583368	BU583368 mai04c04.
	31	16.8	80.0	522	10	BG078216	BG078216 H3025A02-
	32	16.8	80.0	523	13	BQ356299	BQ356299 CM3-HT011
	33	16.8	80.0	526	13	BQ287718	BQ287718 ii68f07.y
	34	16.8	80.0	543	9	AA798972	AA798972 vv95e04.r
	35	16.8	80.0	544	4	BX517147	Bx517147 RZPD Mus
	36	16.8	80.0	554	9	AA543306	AA543306 vk36e11.r
	37	16.8	80.0	556	9	AA543338	AA543338 vk36h11.r
c	38	16.8	80.0	563	14	W19670	W19670 zb36a03.r1
	39	16.8	80.0	581	14	CB271143	CB271143 mai55g11.
	40	16.8	80.0	583	28	AQ047644	AQ047644 cLM-2c9-u
	41	16.8	80.0	587	13	BQ444442	BQ444442 UI-M-ER0-
	42	16.8	80.0	593	13	BQ265035	BQ265035 NISC_ff03
	43	16.8	80.0	600	12	BI987638	BI987638 3204-05 M
	44	16.8	80.0	604	9	AA544735	AA544735 vk35d11.r
c	45	16.8	80.0	640	29	BZ421605	BZ421605 hz32b04.b

ALIGNMENTS

RESULT 1
CC270269
LOCUS CC270269 686 bp DNA linear GSS 13-MAY-2003
DEFINITION CH261-1N21_Sp6.1 CH261 Gallus gallus genomic clone CH261-1N21, genomic survey sequence.
ACCESSION CC270269
VERSION CC270269.1 GI:30623166
KEYWORDS GSS.
SOURCE Gallus gallus (chicken)
ORGANISM Gallus gallus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae; Gallus.
REFERENCE 1 (bases 1 to 686)
AUTHORS Kremitzki,C., Higginbotham,J., Wylie,K., Carter,J., McPherson,J., Warren,W., Graves,T., Mardis,E. and Wilson,R.
TITLE Gallus gallus BAC End Reads
JOURNAL Unpublished
COMMENT Contact: Richard K. Wilson
Genome Sequencing Center
Washington University School of Medicine
Email: submissions@watson.wustl.edu
Insert Length: 182000 Std Error: 0.00
Seq primer: Sp6 ATTTAGGTGACACTATAG
Class: BAC ends
High quality sequence start: 68
High quality sequence stop: 498.

```
FEATURES             Location/Qualifiers
     source          1. .686
                     /organism="Gallus gallus"
                     /mol_type="genomic DNA"
                     /strain="Red Jungle Fowl"
                     /db_xref="taxon:9031"
                     /clone="CH261-1N21"
                     /sex="female"
                     /cell_line="UCD001, inbred 256"
                     /clone_lib="CH261"
                     /note="Vector: pTARBAC2.1; Site_1: EcoRI; Site_2: EcoRI;
                     CH261 Female Chicken library - for library and clone
                     ordering information: http://www.chori.org/bacpac"
BASE COUNT      196 a    147 c    117 g    226 t
ORIGIN

  Query Match             87.6%;  Score 18.4;  DB 29;  Length 686;
  Best Local Similarity   95.0%;  Pred. No. 3.4e+02;
  Matches   19;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          2 AGCCTGATTGATGACCAGGG 21
              |||||||||||| ||||||||
Db        524 AGCCTGATTGAAGACCAGGG 543


RESULT 2
BZ721194
LOCUS       BZ721194                809 bp    DNA     linear   GSS 24-FEB-2003
DEFINITION  PUCFH73TD ZM_0.6_1.0_KB Zea mays genomic clone ZMMBTa133M02,
            genomic survey sequence.
ACCESSION   BZ721194
VERSION     BZ721194.1  GI:28513208
KEYWORDS    GSS.
SOURCE      Zea mays
  ORGANISM  Zea mays
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD
            clade; Panicoideae; Andropogoneae; Zea.
REFERENCE   1  (bases 1 to 809)
  AUTHORS   Whitelaw,C.A., Quackenbush,J., Van Aken,S., Utterback,T., Resnick
            ,A., Fraser,C.M., Yuan,Y., San Miguel,P., Ma,J. and Bennetzen,J.
  TITLE     Maize Genomics Consortium
  JOURNAL   Unpublished
COMMENT     Contact: Cathy Whitelaw
            TIGR
            9712 Medical Center Drive, Rockville, MD 20850, USA
            Tel: 301-838-5843
            Fax: 301-838-0208
            Email: whitelaw@tigr.org
            Seq primer: TF
            Class: sheared ends.
FEATURES             Location/Qualifiers
     source          1. .809
                     /organism="Zea mays"
                     /mol_type="genomic DNA"
                     /strain="B73"
                     /db_xref="taxon:4577"
                     /clone="ZMMBTa133M02"
                     /clone_lib="ZM_0.6_1.0_KB"
                     /note="Vector: pCR4-TOPO; Site_1: EcoRI; 0.6-1.0 kb high
                     CoT selected genomic DNA library"
BASE COUNT      231 a    173 c    172 g    233 t
ORIGIN

  Query Match             85.7%;  Score 18;  DB 29;  Length 809;
  Best Local Similarity   100.0%;  Pred. No. 5.8e+02;
  Matches   18;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          4 CCTGATTGATGACCAGGG 21
              ||||||||||||||||||
Db        773 CCTGATTGATGACCAGGG 790


RESULT 3
AA894450/c
LOCUS       AA894450                307 bp    mRNA    linear   EST 06-APR-1998
DEFINITION  oe79f07.s1 NCI_CGAP_Lu5 Homo sapiens cDNA clone IMAGE:1417861 3',
            mRNA sequence.
ACCESSION   AA894450
VERSION     AA894450.1  GI:3030851
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 307)
  AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
  JOURNAL   Unpublished
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R.
            Emmert-Buck, M.D., Ph.D.
             cDNA Library Preparation: M. Bento Soares, Ph.D.
             cDNA Library Arrayed by: Greg Lennon, Ph.D.
             DNA Sequencing by: Washington University Genome Sequencing Center
             Clone distribution: NCI-CGAP clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            www-bio.llnl.gov/bbrp/image/image.html
            Seq primer: -40m13 fwd. ET from Amersham
            High quality sequence stop: 288.
FEATURES             Location/Qualifiers
     source          1. .307
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:1417861"
                     /tissue_type="carcinoid"
                     /lab_host="DH10B"
                     /clone_lib="NCI_CGAP_Lu5"
                     /note="Organ: lung; Vector: pT7T3D-Pac (Pharmacia) with a
                     modified polylinker; 1st strand cDNA was prepared from
                     neuroendocrine lung carcinoid, and was then primed with a
                     Not I - oligo(dT) primer. Double-stranded cDNA was ligated
                     to Eco RI adaptors (Pharmacia), digested with Not I and
                     cloned into the Not I and Eco RI sites of the modified
                     pT7T3 vector.  Library is normalized. Library was
                     constructed by Bento Soares and M. Fatima Bonaldo. "
BASE COUNT      116 a     62 c     42 g     87 t
ORIGIN

  Query Match             84.8%;  Score 17.8;  DB 9;  Length 307;
  Best Local Similarity   90.5%;  Pred. No. 4.6e+02;
  Matches   19;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy          1 AAGCCTGATTGATGACCAGGG 21
              || |||| ||||||||||||||
Db        247 AAACCTGGTTGATGACCAGGG 227


RESULT 4
AI572687/c
LOCUS       AI572687                339 bp    mRNA    linear   EST 14-MAY-1999
DEFINITION  tr78b10.x1 NCI_CGAP_Pan1 Homo sapiens cDNA clone IMAGE:2224411 3',
            mRNA sequence.
ACCESSION   AI572687
VERSION     AI572687.1  GI:4536061
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 339)
  AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
```

```
            Tumor Gene Index
  JOURNAL   Unpublished
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Life Technologies catalog #: 11548-013
             DNA Sequencing by: Washington University Genome Sequencing Center
             Clone distribution: NCI-CGAP clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            www-bio.llnl.gov/bbrp/image/image.html
            Insert Length: 1905   Std Error: 0.00
            Seq primer: -40UP from Gibco
            High quality sequence stop: 332
            POLYA=No.
FEATURES              Location/Qualifiers
     source           1. .339
                      /organism="Homo sapiens"
                      /mol_type="mRNA"
                      /db_xref="taxon:9606"
                      /clone="IMAGE:2224411"
                      /tissue_type="adenocarcinoma"
                      /lab_host="DH10B"
                      /clone_lib="NCI_CGAP_Pan1"
                      /note="Organ: pancreas; Vector: pCMV-SPORT6; Site_1: SalI;
                      Site_2: NotI; Cloned unidirectionally.  Primer: Oligo dT.
                      Average insert size 1.72 kb. Life Technologies catalog #:
                      11548-013"
BASE COUNT      129 a     67 c     46 g     97 t
ORIGIN

  Query Match             84.8%;  Score 17.8;  DB 9;  Length 339;
  Best Local Similarity   90.5%;  Pred. No. 4.8e+02;
  Matches   19;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy          1 AAGCCTGATTGATGACCAGGG 21
              || |||| ||||||||||||||
Db        253 AAACCTGGTTGATGACCAGGG 233


RESULT 5
BH323391
LOCUS       BH323391                380 bp    DNA     linear   GSS 03-DEC-2001
DEFINITION  CH230-53J16.TJ CHORI-230 Segment 1 Rattus norvegicus genomic clone
            CH230-53J16, genomic survey sequence.
ACCESSION   BH323391
VERSION     BH323391.1  GI:17254105
KEYWORDS    GSS.
SOURCE      Rattus norvegicus (Norway rat)
  ORGANISM  Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Rattus.
REFERENCE   1  (bases 1 to 380)
  AUTHORS   Zhao,S., Shetty,J., Shatsman,S., Tsegaye,G., Geer,K., Shvartsbeyn
            ,A., Gebregeorgis,E., Overton,L., Russell,D., Chen,D., Riggs,F., de
            Jong,P. and Fraser,C.M.
  TITLE     Rat BAC End Sequences from Library CHORI-230 EcoRI segment
  JOURNAL   Unpublished
COMMENT     Other_GSSs: CH230-53J16.TV
            Contact: Shaying Zhao
            Department of Eukaryotic Genomics
            The Institute for Genomic Research
            9712 Medical Center Dr., Rockville, MD 20850, USA
            Tel: 301 838 0200
            Fax: 301 838 0208
            Email: szhao@tigr.org
            Clones are derived from the rat BAC library CHORI-230
            (http://www.chori.org/bacpac/rat230.htm). For BAC library
            availability, please contact Pieter de Jong (pdejong@mail.cho.org).
            Clones may be purchased from BACPAC Resources
            (http://www.chori.org/bacpac/or ering_information.htm). BAC end
            page: http://www.tigr.org/tdb/bac_ends/rat/bac_end_intro.html
            Plate: 53  row: J  column: 16
            Seq primer: SP6
            Class: BAC ends.
FEATURES              Location/Qualifiers
     source           1. .380
                      /organism="Rattus norvegicus"
                      /mol_type="genomic DNA"
                      /strain="BN/SsNHsd/MCW"
                      /db_xref="taxon:10116"
                      /clone="CH230-53J16"
                      /sex="Female"
                      /cell_type="Brain"
                      /clone_lib="CHORI-230 Segment 1"
                      /note="Vector: pTARBAC2.1; Site_1: EcoRI; Site_2: EcoRI;
                      CHORI-230 Rat (BN/SsNHsd/MCW) BAC library produced by
                      Pieter de Jong"
BASE COUNT      105 a     89 c     103 g     83 t
ORIGIN

  Query Match             84.8%;  Score 17.8;  DB 28;  Length 380;
  Best Local Similarity   90.5%;  Pred. No. 5e+02;
  Matches   19;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy          1 AAGCCTGATTGATGACCAGGG 21
              |||||||||||||||| ||| |
Db        147 AAGCCTGATTGATGATCAGAG 167


RESULT 6
CA941354/c
LOCUS       CA941354                545 bp    mRNA    linear   EST 30-DEC-2002
DEFINITION  ir34b01.x1 HR85 islet Homo sapiens cDNA clone IMAGE:6546961 3',
            mRNA sequence.
ACCESSION   CA941354
VERSION     CA941354.1  GI:27429834
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 545)
  AUTHORS   Melton,D., Brown,J., Kenty,G., Permutt,A., Lee,C., Kaestner,K.,
            Lemishka,I., Scearce,M., Brestelli,J., Gradwohl,G., Clifton,S.,
            Hillier,L., Marra,M., Pape,D., Wylie,T., Martin,J., Blistain,A.,
            Schmitt,A., Theising,B., Ritter,E., Ronko,I., Bennett,J., Cardenas
            ,M., Gibbons,M., McCann,R., Cole,R., Tsagareishvili,R., Williams,T.
            , Jackson,Y. and Bowers,Y.
  TITLE     Endocrine Pancreas Consortium
  JOURNAL   Unpublished
COMMENT     Contact: Douglas Melton, Klaus H. Kaestner, & Hiroshi Inoue
            Endocrine Pancreas Consortium
            Harvard University, Howard Hughes Medical Institute
            Dept of Molecular and Cellular Biology, 7 Divinity Ave, Cambridge,
            MA 02138
            Tel: 617-495-1812
            Fax: 617-495-8557
            Email: dmelton@biohp.harvard.edu
            Library was constructed by Dr. Hiroshi Inoue DNA sequencing by:
            Washington University Genome Sequencing Center For information on
            obtaining a clone please contact: Dr. Hiroshi Inoue
            (hinoue@im.wustl.edu)
            Seq primer: -40RP from Gibco
            High quality sequence stop: 452.
FEATURES              Location/Qualifiers
     source           1. .545
                      /organism="Homo sapiens"
                      /mol_type="mRNA"
                      /db_xref="taxon:9606"
                      /clone="IMAGE:6546961"
                      /tissue_type="Purified pancreatic islet"
                      /lab_host="DH10B"
                      /clone_lib="HR85 islet"
                      /note="Organ: Pancreas; Vector: pBluescript SK(-); Site_1:
                      NotI; Site_2: XhoI; cDNA made by oligo-dT priming.
                      Size-selected on agarose gel. Average insert size ~1kb. 5'
```

```
                 XhoI site was destroyed after directional cloning.
                 Amplified once. Contact information: Hiroshi Inoue, MD,
                 Metabolism Div. (Alan Permutt Lab), Washington University
                 School of Medicine, Box 8127, 660 South Euclid Ave., St.
                 Louis, MO 63110, E-mail: hinoue@imgate.wustl.edu, Tel:
                 314-362-1916, Fax: 314-747-2692."
BASE COUNT      184 a    100 c     88 g    173 t
ORIGIN

  Query Match             84.8%;  Score 17.8;  DB 14;  Length 545;
  Best Local Similarity   90.5%;  Pred. No. 6e+02;
  Matches   19;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy          1 AAGCCTGATTGATGACCAGGG 21
              || |||| ||||||||||||||
Db        257 AAACCTGGTTGATGACCAGGG 237


RESULT 7
AI816423/c
LOCUS       AI816423                592 bp    mRNA    linear   EST 09-JUL-1999
DEFINITION  au47g07.x1 Schneider fetal brain 00004 Homo sapiens cDNA clone
            IMAGE:2517948 3', mRNA sequence.
ACCESSION   AI816423
VERSION     AI816423.1  GI:5431969
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 592)
  AUTHORS   Hillier,L., Allen,M., Bowles,L., Dubuque,T., Geisel,G., Jost,S.,
            Krizman,D., Kucaba,T., Lacy,M., Le,N., Lennon,G., Marra,M., Martin
            ,J., Moore,B., Schellenberg,K., Steptoe,M., Tan,F., Theising,B.,
            White,Y., Wylie,T., Waterston,R. and Wilson,R.
  TITLE     WashU-NCI human EST Project
  JOURNAL   Unpublished
COMMENT     Contact: Wilson RK
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: est@watson.wustl.edu
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            Seq primer: -40UP from Gibco
            High quality sequence stop: 469.
FEATURES             Location/Qualifiers
     source          1. .592
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:2517948"
                     /sex="male"
                     /tissue_type="frontal lobe"
                     /dev_stage="5 months post-conception"
                     /lab_host="DH10B"
                     /clone_lib="Schneider fetal brain 00004"
                     /note="Organ: brain; Vector: pBluescript SK (Stratagene);
                     Site_1: SstI; Site_2: XhoI; Double-stranded cDNA was
                     prepared from human fetal brain tissue.  5' and 3'
                     adaptors were used in cloning as follows: 5' adaptor
                     sequence:
                     5'-GAGAGAGAGAAGAGCTCAAGGATCCTTAATTAAATTAATCCCCCCCCCCCCCC-3'
                     and 3' adaptor sequence:
                     5'-GAGAGAGAGACTCGAGTTTTTTTTTTTTTTTTT-3'.  The library was
                     size-selected for >0.5 kb inserts and has an average
                     insert size estimated at 1.2 kb.  This library was
                     constructed using the CAP-trapper method for full-length
                     enrichment and has not undergone amplification.    Library
                     was constructed by Dr. Claudio Schneider (LNCIB-Area
                     Science Park,  Trieste, Italy). "
BASE COUNT      195 a    106 c     98 g    192 t      1 others
ORIGIN

  Query Match             84.8%;  Score 17.8;  DB 9;  Length 592;
  Best Local Similarity   90.5%;  Pred. No. 6.2e+02;
  Matches   19;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy          1 AAGCCTGATTGATGACCAGGG 21
              || |||| ||||||||||||||
Db        259 AAACCTGGTTGATGACCAGGG 239


RESULT 8
AG121419/c
LOCUS       AG121419                671 bp    DNA     linear   GSS 04-NOV-2001
DEFINITION  Pan troglodytes DNA, clone: PTB-130H10.F, genomic survey sequence.
ACCESSION   AG121419
VERSION     AG121419.1  GI:16650584
KEYWORDS    GSS.
SOURCE      Pan troglodytes (chimpanzee)
  ORGANISM  Pan troglodytes
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Pan.
REFERENCE   1
  AUTHORS   Fujiyama,A., Hattori,M., Toyoda,A., Taylor,T.D., Yada,T.,
            Totoki,Y., Watanabe,H. and Sakaki,Y.
  TITLE     BAC end sequences of Library PTB
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 671)
  AUTHORS   Fujiyama,A., Hattori,M., Toyoda,A., Taylor,T.D., Yada,T.,
            Totoki,Y., Watanabe,H. and Sakaki,Y.
  TITLE     Direct Submission
  JOURNAL   Submitted (02-AUG-2001) Asao Fujiyama, The Institute of Physical
            and Chemical Research (RIKEN), Genomic Sciences Center (GSC);
            1-7-22 Suehiro-chou,Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
            (E-mail:chimpbes@gsc.riken.go.jp, URL:http://hgp.gsc.riken.go.jp/,
            Tel:81-45-503-9111, Fax:81-45-503-9170)
COMMENT     Clones are derived from the chimpanzee BAC library PTB This BAC end
            was generated during the R&D process and may have higher chance of
            clone tracking errors.
             PRIMERS
                 Sequencing: -21M13
             LIBRARY
                 Vector    : pKS145
                 R.Site 1  : SacI
                 R.Site 2  : SacI.
FEATURES             Location/Qualifiers
     source          1. .671
                     /organism="Pan troglodytes"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9598"
                     /clone="PTB-130H10.F"
                     /sex="male"
                     /cell_type="lymphoblast"
                     /clone_lib="PTB Chimpanzee Male BAC Library"
BASE COUNT      195 a    161 c     152 g    163 t
ORIGIN

  Query Match             84.8%;  Score 17.8;  DB 29;  Length 671;
  Best Local Similarity   90.5%;  Pred. No. 6.6e+02;
  Matches   19;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy          1 AAGCCTGATTGATGACCAGGG 21
              |||||||||| || |||||||
Db        414 AAGCCTGATTTATCACCAGGG 394


RESULT 9
AG186976/c
LOCUS       AG186976                708 bp    DNA     linear   GSS 05-JAN-2002
DEFINITION  Pan troglodytes DNA, clone:PTB-004E01.R, genomic survey sequence.
ACCESSION   AG186976
VERSION     AG186976.1  GI:18149509
KEYWORDS    GSS.
```

```
SOURCE      Pan troglodytes (chimpanzee)
  ORGANISM  Pan troglodytes
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Pan.
REFERENCE   1
  AUTHORS   Fujiyama,A., Hattori,M., Toyoda,A., Taylor,T.D., Yada,T.,
            Totoki,Y., Watanabe,H. and Sakaki,Y.
  TITLE     BAC end sequences of Library PTB
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 708)
  AUTHORS   Fujiyama,A., Hattori,M., Toyoda,A., Taylor,T.D., Yada,T.,
            Totoki,Y., Watanabe,H. and Sakaki,Y.
  TITLE     Direct Submission
  JOURNAL   Submitted (05-DEC-2001) Asao Fujiyama, The Institute of Physical
            and Chemical Research (RIKEN), Genomic Sciences Center (GSC);
            1-7-22 Suehiro-chou,Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
            (E-mail:chimpbes@gsc.riken.go.jp, URL:http://hgp.gsc.riken.go.jp/,
            Tel:81-45-503-9111, Fax:81-45-503-9170)
COMMENT     Clones are derived from the chimpanzee BAC library PTB This BAC end
            was generated during the R&D process and may have higher chance of
            clone tracking errors.
             PRIMERS
                Sequencing: M13Rev
             LIBRARY
                Vector    : pKS145
                R.Site 1  : SacI
                R.Site 2  : SacI.
FEATURES             Location/Qualifiers
     source          1. .708
                     /organism="Pan troglodytes"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9598"
                     /clone="PTB-004E01.R"
                     /sex="male"
                     /cell_type="lymphoblast"
                     /clone_lib="PTB Chimpanzee Male BAC Library"
BASE COUNT      201 a    166 c    156 g    182 t      3 others
ORIGIN

  Query Match             84.8%;  Score 17.8;  DB 29;  Length 708;
  Best Local Similarity   90.5%;  Pred. No. 6.7e+02;
  Matches   19;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy          1 AAGCCTGATTGATGACCAGGG 21
              ||||||||||  ||  |||||||
Db        398 AAGCCTGATTTATCACCAGGG 378


RESULT 10
BU245702
LOCUS       BU245702                 913 bp    mRNA    linear   EST 26-NOV-2002
DEFINITION  603783316F1 CSEQCHN34 Gallus gallus cDNA clone ChEST735o4 5', mRNA
            sequence.
ACCESSION   BU245702
VERSION     BU245702.1  GI:25493069
KEYWORDS    EST.
SOURCE      Gallus gallus (chicken)
  ORGANISM  Gallus gallus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Archosauria; Aves; Neognathae; Galliformes; Phasianidae;
            Phasianinae; Gallus.
REFERENCE   1  (bases 1 to 913)
  AUTHORS   Boardman,P.E., Sanz-Ezquerro,J., Overton,I.M., Burt,D.W., Bosch,E.,
            Fong,W.T., Tickle,C., Brown,W.R.A., Wilson,S.A. and Hubbard,S.J.
  TITLE     A Comprehensive Collection of Chicken cDNAs
  JOURNAL   Curr. Biol. 12 (22), 1965-1969 (2002)
  MEDLINE   22335534
   PUBMED   12445392
COMMENT     Contact: Simon Hubbard
            Department of Biomolecular Sciences
            University of Manchester Institute of Science and Technology (UMIST
            )
            PO Box 88, Manchester, M60 1QD, UK
            Tel: 01612008930
            Fax: 01612360409
            Email: Simon.Hubbard@umist.ac.uk.
FEATURES             Location/Qualifiers
     source          1. .913
                     /organism="Gallus gallus"
                     /mol_type="mRNA"
                     /strain="Compton Line 15I"
                     /db_xref="taxon:9031"
                     /clone="ChEST735o4"
                     /sex="Female"
                     /dev_stage="adult"
                     /lab_host="DH10B"
                     /clone_lib="CSEQCHN34"
                     /note="Organ: liver; Vector: pBluescript II KS(+); Site_1:
                     EcoRI; Site_2: NotI; This normalized library was
                     constructed from 1 million independent clones. cDNA
                     synthesis was initiated using an oligo(dT) primer, using
                     methylated C in the first strand synthesis reaction.
                     Following this first strand reaction, double-stranded cDNA
                     was blunted, ligated to NotI adapters, digested with EcoRI
                     , size-selected, and cloned into the NotI and EcoRI
                     compatible sites of a custom modified MCS of the
                     pBluescript (KS+) vector. The library was normalized in 2
                     rounds using conditions adapted from Soares et al., PNAS
                     (1994) 91: 9228-9232 and Bonaldo et al., Genome Research 6
                     (1996): 791, except that a significantly longer
                     reannealing hybridization was used."
BASE COUNT      217 a    230 c    207 g    259 t
ORIGIN

  Query Match             84.8%;  Score 17.8;  DB 13;  Length 913;
  Best Local Similarity   90.5%;  Pred. No. 7.6e+02;
  Matches   19;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy          1 AAGCCTGATTGATGACCAGGG 21
              ||||||||| |||| ||||||
Db        145 AAGCCTGATGGATGGCCAGGG 165


RESULT 11
BG119393
LOCUS       BG119393                 963 bp    mRNA    linear   EST 30-JAN-2001
DEFINITION  602349076F1 NIH_MGC_90 Homo sapiens cDNA clone IMAGE:4444074 5',
            mRNA sequence.
ACCESSION   BG119393
VERSION     BG119393.1  GI:12612899
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 963)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: ATCC
             cDNA Library Preparation: Life Technologies, Inc.
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Incyte Genomics, Inc.
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM10219  row: g  column: 19
            High quality sequence stop: 658.
FEATURES             Location/Qualifiers
     source          1. .963
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:4444074"
```

```
                     /tissue_type="adenocarcinoma, cell line"
                     /lab_host="DH10B (phage-resistant)"
                     /clone_lib="NIH_MGC_90"
                     /note="Organ: liver; Vector: pCMV-SPORT6; Site_1: NotI;
                     Site_2: SalI; Cloned unidirectionally; oligo-dT primed.
                     Average insert size 1.7 kb. Library enriched for
                     full-length clones and constructed by Life Technologies.
                     Note: this is a NIH_MGC Library."
BASE COUNT      285 a    199 c    199 g    280 t
ORIGIN

  Query Match             84.8%;  Score 17.8;  DB 10;  Length 963;
  Best Local Similarity   90.5%;  Pred. No. 7.8e+02;
  Matches   19;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy          1 AAGCCTGATTGATGACCAGGG 21
              || |||| ||||||||||||||
Db        570 AAACCTGGTTGATGACCAGGG 590


RESULT 12
AL523235/c
LOCUS       AL523235                1201 bp    mRNA    linear   EST 22-MAY-2003
DEFINITION  AL523235 Homo sapiens NEUROBLASTOMA COT 25-NORMALIZED Homo sapiens
            cDNA clone CS0DC001YE18 3-PRIME, mRNA sequence.
ACCESSION   AL523235
VERSION     AL523235.2  GI:31041496
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 1201)
  AUTHORS   Li,W.B., Gruber,C., Jessee,J. and Polayes,D.
  TITLE     Full-length cDNA libraries and normalization
  JOURNAL   Unpublished
COMMENT     On Feb 13, 2001 this sequence version replaced gi:12786728.
            Contact: Genoscope
            Genoscope - Centre National de Sequencage
            BP 191 91006 EVRY cedex - France
            Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr
            Library was constructed by Life Technologies, a division of
            Invitrogen. This sequence belongs to sequence cluster 10668.r, and
            it belongs to a clone representative of this cluster. For more
            information about this cluster and the virtual cDNA, see
            http://www.genoscope.cns.fr/
            cgi-bin/cluster.cgi?seq=CS0DC001BC09NP1&cluster=10668.r. Contact :
            Feng Liang Email : fliang@lifetech.com URL :
            http://fulllength.invitrogen.com/ InVitroGen Corporation 1600
            Faraday Avenue Genoscope sequence ID : CS0DC001BC09NP1.
FEATURES             Location/Qualifiers
     source          1. .1201
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="CS0DC001YE18"
                     /tissue_type="NEUROBLASTOMA COT 25-NORMALIZED"
                     /clone_lib="Homo sapiens NEUROBLASTOMA COT 25-NORMALIZED"
                     /note="1st strand cDNA was primed with a NotI-oligo(dT)
                     primer. Five prime end enriched, double-strand cDNA was
                     digested with Not I and cloned into the Not I and EcoR V
                     sites of the pCMVSPORT 6 vector. Library was normalized."
BASE COUNT      358 a    206 c    252 g    311 t     74 others
ORIGIN

  Query Match             84.8%;  Score 17.8;  DB 9;  Length 1201;
  Best Local Similarity   90.5%;  Pred. No. 8.6e+02;
  Matches   19;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy          1 AAGCCTGATTGATGACCAGGG 21
              || |||| ||||||||||||||
Db        236 AAACCTGGTTGATGACCAGGG 216


RESULT 13
AZ072191
LOCUS       AZ072191                565 bp    DNA     linear   GSS 31-MAR-2000
DEFINITION  RPCI-23-396I8.TV RPCI-23 Mus musculus genomic clone RPCI-23-396I8,
            genomic survey sequence.
ACCESSION   AZ072191
VERSION     AZ072191.1  GI:7365088
KEYWORDS    GSS.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 565)
  AUTHORS   Zhao,S., Nierman,W., Feldblyum,T., Malek,J., Shatsman,S., Akinret
            ,B., Levins,M., Mcgann,S., Tsegaye,G., Geer,K., Krol,M., de Jong,P.
            and Fraser,C.M.
  TITLE     Mouse BAC End Sequences from Library RPCI-23
  JOURNAL   Unpublished
COMMENT     Contact: Shaying Zhao
            Department of Eukaryotic Genomics
            The Institute for Genomic Research
            9712 Medical Center Dr., Rockville, MD 20850, USA
            Tel: 301 838 0200
            Fax: 301 838 0208
            Email: szhao@tigr.org
            Clones are derived from the mouse BAC library RPCI-23. For BAC
            library availability, please contact Pieter de Jong
            (pieter@dejong.med.buffalo.edu). Clones may be purchased from
            BACPAC Resources (http://bacpac.med.buffalo.edu/orderingframe.htm)
            or from Resea ch Genetics (info@resgen.com). BAC end page:
            http://www.tigr.org/tdb/bac_ends/mouse/bac_end_intro.html
            Plate: 396  row: I  column: 8
            Seq primer: T7
            Class: BAC ends.
FEATURES             Location/Qualifiers
     source          1. .565
                     /organism="Mus musculus"
                     /mol_type="genomic DNA"
                     /strain="C57BL/6J"
                     /db_xref="taxon:10090"
                     /clone="RPCI-23-396I8"
                     /sex="Female"
                     /lab_host="DH10B"
                     /clone_lib="RPCI-23"
                     /note="Organ: Kidney/Brain; Vector: pBACe3.6; Site_1:
                     EcoRI; Site_2: EcoRI; Female C57BL/6J mouse kidney and/or
                     brain genomic DNA was isolated and  partially digested
                     with a combination of EcoRI and EcoRI Methylase.   Size
                     selected DNA was cloned into the pBACe3.6 vector at the
                     EcoRI  sites.  The ligation products were transformed into
                     DH10B  electrocompetent cells (BRL Life Technologies).   "
BASE COUNT      142 a    131 c    162 g    130 t
ORIGIN

  Query Match             82.9%;  Score 17.4;  DB 28;  Length 565;
  Best Local Similarity   94.7%;  Pred. No. 9.4e+02;
  Matches   18;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          2 AGCCTGATTGATGACCAGG 20
              |||||||||||||| |||||
Db         26 AGCCTGATTGATGCCCAGG 44


RESULT 14
BE729033/c
LOCUS       BE729033                914 bp    mRNA    linear   EST 15-SEP-2000
DEFINITION  601561212F1 NIH_MGC_20 Homo sapiens cDNA clone IMAGE:3830835 5',
            mRNA sequence.
ACCESSION   BE729033
VERSION     BE729033.1  GI:10143025
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
```

```
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 914)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: ATCC/DCTD/DTP
             cDNA Library Preparation: Ling Hong/Rubin Laboratory
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Incyte Genomics, Inc.
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at: image.llnl.gov
            Plate: LLCM505  row: h  column: 04
            High quality sequence stop: 694.
FEATURES              Location/Qualifiers
     source           1. .914
                      /organism="Homo sapiens"
                      /mol_type="mRNA"
                      /db_xref="taxon:9606"
                      /clone="IMAGE:3830835"
                      /tissue_type="melanotic melanoma"
                      /lab_host="DH10B (phage-resistant)"
                      /clone_lib="NIH_MGC_20"
                      /note="Organ: skin; Vector: pOTB7; Site_1: XhoI; Site_2:
                      EcoRI; cDNA made by oligo-dT priming.  Directionally
                      cloned into EcoRI/XhoI sites using the following 5'
                      adaptor: GGCACGAG(G).  Size-selected >500bp for average
                      insert size 1.8kb.  Library constructed by Ling Hong in
                      the laboratory of Gerald M. Rubin (University of
                      California, Berkeley) using  ZAP-cDNA synthesis kit
                      (Stratagene) and Superscript II RT (Life Technologies)."
BASE COUNT      212 a    241 c    238 g    223 t
ORIGIN

  Query Match             82.9%;  Score 17.4;  DB 10;  Length 914;
  Best Local Similarity   94.7%;  Pred. No. 1.2e+03;
  Matches   18;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          1 AAGCCTGATTGATGACCAG 19
              ||||||||||| ||||||
Db        626 AAGCCTGATTGAGGACCAG 608


RESULT 15
AI818091/c
LOCUS       AI818091                 543 bp    mRNA    linear   EST 21-DEC-1999
DEFINITION  wk27a09.x1 NCI_CGAP_Brn25 Homo sapiens cDNA clone IMAGE:2413528 3'
            similar to SW:HIS1_BACSU O34520 ATP PHOSPHORIBOSYLTRANSFERASE ;,
            mRNA sequence.
ACCESSION   AI818091
VERSION     AI818091.1  GI:5437170
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 543)
  AUTHORS   NCI/NINDS-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE     National Cancer Institute / National Institute of Neurological
            Disorders and Stroke, Brain Tumor Genome Anatomy Project
            (CGAP/BTGAP), Tumor Gene Index
  JOURNAL   Unpublished
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: David N. Louis, M.D., Myrna R. Rosenfeld M.D.,
            Ph.D.
             cDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima
            Bonaldo, Ph.D.
             cDNA Library Arrayed by: Greg Lennon, Ph.D.
             DNA Sequencing by: Washington University Genome Sequencing Center
             Clone distribution: NCI-CGAP clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            www-bio.llnl.gov/bbrp/image/image.html
            Insert Length: 904    Std Error: 0.00
            Seq primer: -40UP from Gibco
            High quality sequence stop: 485.
FEATURES              Location/Qualifiers
     source           1. .543
                      /organism="Homo sapiens"
                      /mol_type="mRNA"
                      /db_xref="taxon:9606"
                      /clone="IMAGE:2413528"
                      /tissue_type="anaplastic oligodendroglioma"
                      /lab_host="DH10B"
                      /clone_lib="NCI_CGAP_Brn25"
                      /note="Organ: brain; Vector: pT7T3D-Pac (Pharmacia) with a
                      modified polylinker; Site_1: Not I; Site_2: Eco RI; 1st
                      strand cDNA was primed with a Not I - oligo(dT) primer [5'
                      TGTTACCAATCTGAAGTGGGAGCGGCCGCATAGGTTTTTTTTTTTTTTTTTTTTTT
                      T 3']; double-stranded cDNA was ligated to Eco RI
                      adaptors (Pharmacia), digested with Not I and cloned into
                      the Not I and Eco RI sites of  the modified pT7T3 vector.
                      Library is normalized, and was constructed  by Bento
                      Soares and M.Fatima Bonaldo."
BASE COUNT       99 a    174 c    180 g     90 t
ORIGIN

  Query Match             81.0%;  Score 17;  DB 9;  Length 543;
  Best Local Similarity   100.0%;  Pred. No. 1.4e+03;
  Matches   17;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          3 GCCTGATTGATGACCAG 19
              |||||||||||||||||
Db        494 GCCTGATTGATGACCAG 478


Search completed: February 19, 2004, 23:14:36
Job time : 1575 secs
```

GenCore version 5.1.6

OM nucleic - nucleic search, using sw model

Run on: February 19, 2004, 19:59:22 ; Search time 175.5 Seconds
(without alignments)
323.010 Million cell updates/sec

Title: US-10-085-108-21_COPY_711_731
Perfect score: 21
Sequence: 1 AAGCCTGATTGATGACCAGGG 21

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 2552756 seqs, 1349719017 residues

Total number of hits satisfying chosen parameters: 5105512

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N_Geneseq_19Jun03:*
1: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1980.DAT:*
2: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1981.DAT:*
3: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1982.DAT:*
4: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1983.DAT:*
5: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1984.DAT:*
6: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1985.DAT:*
7: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1986.DAT:*
8: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1987.DAT:*
9: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1988.DAT:*
10: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1989.DAT:*
11: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1990.DAT:*
12: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1991.DAT:*
13: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1992.DAT:*
14: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1993.DAT:*
15: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1994.DAT:*
16: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1995.DAT:*
17: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1996.DAT:*
18: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1997.DAT:*
19: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1998.DAT:*
20: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1999.DAT:*
21: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA2000.DAT:*
22: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA2001A.DAT:*
23: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA2001B.DAT:*
24: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA2002.DAT:*
25: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA2003.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.		Score	% Query Match	Length	DB	ID	Description
c	1	21	100.0	21	25	ABX95022	Human MAGE-C3 expr
c	2	21	100.0	425	22	ABA46075	Human breast cell
c	3	21	100.0	425	22	ABA56617	Human foetal liver
c	4	21	100.0	425	22	ABA26230	Probe #4696 for ge
c	5	21	100.0	425	22	AAK04758	Human brain expres
c	6	21	100.0	425	22	AAK30282	Human bone marrow
c	7	21	100.0	425	22	AAI14882	Probe #4815 for ge
c	8	21	100.0	425	22	AAI36237	Probe #4923 used t
c	9	21	100.0	425	22	AAI04666	Probe #4657 used t
c	10	21	100.0	425	23	ABS29927	Human liver single
c	11	21	100.0	425	24	ABS04866	Human genome-deriv
	12	21	100.0	1041	25	ABX95006	cDNA encoding huma
	13	21	100.0	7806	23	AAS88354	DNA encoding novel
	14	17	81.0	336	16	AAT24776	Human gene signatu
	15	16.8	80.0	504	24	ABS69805	Novel murine polyn
c	16	16.2	77.1	202	19	AAX10827	Human biallelic po
	17	16.2	77.1	419	23	ABV17420	Human prostate exp
	18	16.2	77.1	456	22	ABA58928	Human foetal liver
	19	16.2	77.1	456	22	AAK07090	Human brain expres
	20	16.2	77.1	456	22	AAK32830	Human bone marrow
	21	16.2	77.1	456	22	AAI38645	Probe #7331 used t
	22	16.2	77.1	456	23	ABS32555	Human liver single
	23	16.2	77.1	456	24	ABS07634	Human genome-deriv
	24	16.2	77.1	458	23	ABV47214	Human prostate exp
	25	16.2	77.1	964	22	ABA15191	Human nervous syst
c	26	16.2	77.1	1236	23	ABL03601	Drosophila melanog
c	27	16.2	77.1	1781	22	ABA18978	Human nervous syst
	28	16.2	77.1	1920	22	AAD10827	Manihot esculenta
	29	16.2	77.1	2204	22	ABA15194	Human nervous syst
	30	16.2	77.1	3171	22	AAC85455	Cold-active beta g
c	31	16.2	77.1	3479	23	ABL03600	Drosophila melanog
c	32	16.2	77.1	3968	23	ABL03602	Drosophila melanog
c	33	16.2	77.1	4912	21	AAA65349	NDO related comple
c	34	16.2	77.1	6779	21	AAA65350	NDO related comple
	35	16.2	77.1	7325	24	AAS94817	Human DNA sequence
c	36	15.8	75.2	470	22	ABA26707	Probe #5173 for ge
c	37	15.8	75.2	1378	23	ABL07577	Drosophila melanog
c	38	15.8	75.2	3116	23	ABL03310	Drosophila melanog
	39	15.8	75.2	3378	23	ABL07576	Drosophila melanog
	40	15.8	75.2	4764	23	ABL05675	Drosophila melanog
	41	15.8	75.2	6147	22	ABA14804	Human nervous syst
	42	15.8	75.2	7300	23	ABL05674	Drosophila melanog
	43	15.4	73.3	312	22	AAH66909	C glutamicum codin
	44	15.4	73.3	320	22	AAK87704	Human digestive sy
	45	15.4	73.3	320	22	AAS31687	Human liver associ

ALIGNMENTS

RESULT 1
ABX95022/c
ID ABX95022 standard; DNA; 21 BP.
XX
AC ABX95022;
XX
DT 05-JUN-2003 (first entry)
XX
DE Human MAGE-C3 expression pattern anlaysis RT-PCR antisense primer.
XX
KW TRAP; ss; tumour rejection antigen precursor; cytolytic T-cell; CTL;
KW tumour; seminoma; bladder transitional-cell carcinoma; NSCLC; adaptor;
KW head-and-neck squamous-cell carcinoma; breast carcinoma; sarcoma;
KW cutaneous melanoma; nonsmall cell lung cancer; RT-PCR; primer; MAGE-C3;
KW human; reverse transcription.
XX
OS Homo sapiens.
XX
PN US2002176865-A1.
XX
PD 28-NOV-2002.
XX
PF 01-MAR-2002; 2002US-0085108.
XX
PR 09-FEB-2000; 2000US-0501104.
PR 25-APR-1997; 97US-0845528.
PR 24-APR-1998; 98US-0066281.
PR 17-DEC-1999; 99US-0468433.
XX
PA (LUCA/) LUCAS S.
PA (BOON/) BOON-FALLEUR T.

XX
PI Lucas S, Boon-Falleur T;
XX
DR WPI; 2003-328468/31.
XX
PT Novel isolated nucleic acid encoding tumor rejection antigen precursor
PT MAGE-C3, MAGE-B5, or MAGE-B6, useful as diagnostic probes to determine
PT presence of abnormal e.g., tumor cells expressing MAGE-C1, MAGE-B5 or
PT MAGE-B6 -
XX
PS Example 12; Page 13; 59pp; English.
XX
CC The invention relates to an isolated nucleic acid molecule which encodes
CC a tumour rejection antigen precursor (TRAP) having an amino acid sequence
CC of a TRAP encoded by a fully defined MAGE-C3, MAGE-B5, or MAGE-B6
CC polynucleotide sequence. Also disclosed is a method which is useful for
CC determining presence of cytolytic T-cells specific for complexes of human
CC leukocyte antigen (HLA) and a peptide derived from the nucleic acid in a
CC cytotoxic T-lymphocyte (CTL)-containing sample. The nucleic acid is
CC useful as a diagnostic probe to determine the presence of abnormal
CC (tumour) cells such as seminoma, bladder transitional-cell carcinoma,
CC head-and-neck squamous-cell carcinoma, breast carcinoma, sarcoma,
CC cutaneous melanoma or nonsmall cell lung cancer (NSCLC) which express
CC MAGE-C1, MAGE-B5 or MAGE-B6. The nucleic acid is useful for diagnosing a
CC disorder characterised by expression of MAGE-C1, MAGE-B5 or MAGE-B6 TRAPs
CC or tumour rejection antigens (TRAs). The present sequence represents the
CC human MAGE-C3 expression pattern anlaysis reverse transcription (RT)-PCR
CC antisense primer.
XX
SQ Sequence 21 BP; 4 A; 7 C; 4 G; 6 T; 0 other;

Query Match 100.0%; Score 21; DB 25; Length 21;
Best Local Similarity 100.0%; Pred. No. 0.39;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 AAGCCTGATTGATGACCAGGG 21
||||||||||||||||||||
Db 21 AAGCCTGATTGATGACCAGGG 1

RESULT 2
ABA46075/c
ID ABA46075 standard; DNA; 425 BP.
XX
AC ABA46075;
XX
DT 01-FEB-2002 (first entry)
XX
DE Human breast cell single exon nucleic acid probe #4770.
XX
KW Human; microarray; single exon probe; gene expression; breast;
KW disease; cancer; ss.
XX
OS Homo sapiens.
XX
PN WO200157271-A2.
XX
PD 09-AUG-2001.
XX
PF 30-JAN-2001; 2001WO-US00662.
XX
PR 04-FEB-2000; 2000US-0180312.
PR 26-MAY-2000; 2000US-0207456.
PR 30-JUN-2000; 2000US-0608408.
PR 03-AUG-2000; 2000US-0632366.
PR 21-SEP-2000; 2000US-0234687.
PR 27-SEP-2000; 2000US-0236359.
PR 04-OCT-2000; 2000GB-0024263.
XX
PA (MOLE-) MOLECULAR DYNAMICS INC.
XX
PI Penn SG, Hanzel DK, Chen W, Rank DR;
XX
DR WPI; 2001-496933/54.
XX
PT New spatially-addressable set of single exon nucleic acid probes,
PT useful for measuring gene expression in sample derived from human
PT breast, comprises number of single exon nucleic acid probes -
XX
PS Claim 1; SEQ ID NO 4770; 327pp + sequence listing; English.
XX
CC The invention relates to a spatially-addressable set of single exon
CC nucleic acid probes for measuring gene expression in a sample derived
CC from human breast and BT 474 cells. The method involves contacting
CC the probes with a collection of detectably labelled nucleic acids
CC derived from mRNA of human breast, and then measuring the label
CC bound to each probe of the microarray. The probes are useful for
CC verifying the expression of regions of genomic DNA predicted to
CC encode proteins. They are useful for gene discovery, and for
CC determining predisposition and/or prognosing breast disease. Gene
CC expression analysis is useful for assessing the toxicity of chemical
CC agents on cells. The microarray of this invention presents a far greater
CC diversity of probes for measuring gene expression, with far less bias
CC than expressed sequence tag microarrays. The method is suitable for
CC rapid production of functional information from genomic sequence. The
CC present sequence is a single exon nucleic acid probe of the invention.
CC Note: The sequence data for this patent did not form part of the
CC printed specification, but was obtained in electronic format directly
CC from WIPO at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 425 BP; 112 A; 107 C; 94 G; 112 T; 0 other;

Query Match 100.0%; Score 21; DB 22; Length 425;
Best Local Similarity 100.0%; Pred. No. 0.69;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 AAGCCTGATTGATGACCAGGG 21
||||||||||||||||||||
Db 150 AAGCCTGATTGATGACCAGGG 130

RESULT 3
ABA56617/c
ID ABA56617 standard; DNA; 425 BP.
XX
AC ABA56617;
XX
DT 01-FEB-2002 (first entry)
XX
DE Human foetal liver single exon nucleic acid probe #4922.
XX
KW Human; foetal liver; gene expression; single exon nucleic acid probe; ss.
XX
OS Homo sapiens.
XX
PN WO200157277-A2.
XX
PD 09-AUG-2001.
XX
PF 30-JAN-2001; 2001WO-US00669.
XX
PR 04-FEB-2000; 2000US-0180312.
PR 26-MAY-2000; 2000US-0207456.
PR 30-JUN-2000; 2000US-0608408.
PR 03-AUG-2000; 2000US-0632366.
PR 21-SEP-2000; 2000US-0234687.
PR 27-SEP-2000; 2000US-0236359.
PR 04-OCT-2000; 2000GB-0024263.
XX
PA (MOLE-) MOLECULAR DYNAMICS INC.
XX
PI Penn SG, Hanzel DK, Chen W, Rank DR;
XX
DR WPI; 2001-483447/52.
XX
PT Human genome-derived single exon nucleic acid probes useful for

PT analyzing gene expression in human fetal liver -
XX
PS Claim 1; SEQ ID NO 4922; 639pp + sequence listing; English.
XX
CC The invention relates to a single exon nucleic acid probe for
CC measuring human gene expression in a sample derived from human foetal
CC liver. The single exon nucleic acid probes may be used for predicting,
CC measuring and displaying gene expression in samples derived from human
CC fetal liver. The present sequence is a single exon nucleic acid
CC probe of the invention.
CC Note: The sequence data for this patent did not form part of the
CC printed specification, but was obtained in electronic format directly
CC from WIPO at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 425 BP; 112 A; 107 C; 94 G; 112 T; 0 other;

Query Match 100.0%; Score 21; DB 22; Length 425;
Best Local Similarity 100.0%; Pred. No. 0.69;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 AAGCCTGATTGATGACCAGGG 21
||||||||||||||||||||
Db 150 AAGCCTGATTGATGACCAGGG 130

RESULT 4
ABA26230/c
ID ABA26230 standard; DNA; 425 BP.
XX
AC ABA26230;
XX
DT 23-JAN-2002 (first entry)
XX
DE Probe #4696 for gene expression analysis in human heart cell sample.
XX
KW Human; gene expression; heart; microarray; vascular system; probe;
KW cardiovascular disease; hypertension; cardiac arrhythmia;
KW congenital heart disease; ss.
XX
OS Homo sapiens.
XX
PN WO200157274-A2.
XX
PD 09-AUG-2001.
XX
PF 30-JAN-2001; 2001WO-US00666.
XX
PR 04-FEB-2000; 2000US-0180312.
PR 26-MAY-2000; 2000US-0207456.
PR 30-JUN-2000; 2000US-0608408.
PR 03-AUG-2000; 2000US-0632366.
PR 21-SEP-2000; 2000US-0234687.
PR 27-SEP-2000; 2000US-0236359.
PR 04-OCT-2000; 2000GB-0024263.
XX
PA (MOLE-) MOLECULAR DYNAMICS INC.
XX
PI Penn SG, Hanzel DK, Chen W, Rank DR;
XX
DR WPI; 2001-488899/53.
XX
PT Single exon nucleic acid probes for analyzing gene expression in human
PT hearts -
XX
PS Claim 1; SEQ ID No 4696; 530pp; English.
XX
CC The present invention relates to single exon nucleic acid probes for
CC measuring human gene expression in a sample derived from human heart. The
CC present sequence is one such probe. The probes may be used for
CC predicting, measuring and displaying gene expression in samples derived
CC from the human heart via microarrays. By measuring gene expression, the
CC probes are useful for predicting, diagnosing, grading, staging,
CC monitoring and prognosing diseases of the human heart and vascular system
CC e.g. cardiovascular disease, hypertension, cardiac arrhythmias and
CC congenital heart disease.
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 425 BP; 112 A; 107 C; 94 G; 112 T; 0 other;

Query Match 100.0%; Score 21; DB 22; Length 425;
Best Local Similarity 100.0%; Pred. No. 0.69;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 AAGCCTGATTGATGACCAGGG 21
||||||||||||||||||||
Db 150 AAGCCTGATTGATGACCAGGG 130

RESULT 5
AAK04758/c
ID AAK04758 standard; DNA; 425 BP.
XX
AC AAK04758;
XX
DT 05-NOV-2001 (first entry)
XX
DE Human brain expressed single exon probe SEQ ID NO: 4749.
XX
KW Human; brain expressed exon; gene expression analysis; probe;
KW microarray; Alzheimer's disease; multiple sclerosis; schizophrenia;
KW epilepsy; cancer; ss.
XX
OS Homo sapiens.
XX
PN WO200157275-A2.
XX
PD 09-AUG-2001.
XX
PF 30-JAN-2001; 2001WO-US00667.
XX
PR 04-FEB-2000; 2000US-0180312.
PR 26-MAY-2000; 2000US-0207456.
PR 30-JUN-2000; 2000US-0608408.
PR 03-AUG-2000; 2000US-0632366.
PR 21-SEP-2000; 2000US-0234687.
PR 27-SEP-2000; 2000US-0236359.
PR 04-OCT-2000; 2000GB-0024263.
XX
PA (MOLE-) MOLECULAR DYNAMICS INC.
XX
PI Penn SG, Hanzel DK, Chen W, Rank DR;
XX
DR WPI; 2001-483446/52.
XX
PT Single exon nucleic acid probes for analyzing gene expression in human
PT brains -
XX
PS Example 4; SEQ ID NO: 4749; 650pp + Sequence Listing; English.
XX
CC The present invention provides a number of single exon nucleic acid
CC probes which are derived from genomic sequences expressed in the human
CC brain. They can be used to measure gene expression in brain cell samples,
CC which may enable the diagnosis and improved treatment of nervous system
CC diseases such as Alzheimer's disease, multiple sclerosis, schizophrenia,
CC epilepsy and cancers. The present sequence is one of the probes of the
CC invention.
XX
SQ Sequence 425 BP; 112 A; 107 C; 94 G; 112 T; 0 other;

Query Match 100.0%; Score 21; DB 22; Length 425;
Best Local Similarity 100.0%; Pred. No. 0.69;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 AAGCCTGATTGATGACCAGGG 21

```
                    ||||||||||||||||||||
Db        150 AAGCCTGATTGATGACCAGGG 130


RESULT 6
AAK30282/c
ID   AAK30282 standard; DNA; 425 BP.
XX
AC   AAK30282;
XX
DT   06-NOV-2001  (first entry)
XX
DE   Human bone marrow expressed single exon probe SEQ ID NO: 4839.
XX
KW   Human; bone marrow expressed exon; gene expression analysis; probe;
KW   microarray; cancer; leukaemia; lymphoma; myeloma; ss.
XX
OS   Homo sapiens.
XX
PN   WO200157276-A2.
XX
PD   09-AUG-2001.
XX
PF   30-JAN-2001; 2001WO-US00668.
XX
PR   04-FEB-2000; 2000US-0180312.
PR   26-MAY-2000; 2000US-0207456.
PR   30-JUN-2000; 2000US-0608408.
PR   03-AUG-2000; 2000US-0632366.
PR   21-SEP-2000; 2000US-0234687.
PR   27-SEP-2000; 2000US-0236359.
PR   04-OCT-2000; 2000GB-0024263.
XX
PA   (MOLE-) MOLECULAR DYNAMICS INC.
XX
PI   Penn SG,  Hanzel DK,  Chen W,  Rank DR;
XX
DR   WPI; 2001-488900/53.
XX
PT   Human genome-derived single exon nucleic acid probes useful for
PT   analyzing gene expression in human bone marrow -
XX
PS   Example 4; SEQ ID NO: 4839; 658pp + Sequence Listing; English.
XX
CC   The present invention provides a number of single exon nucleic acid
CC   probes which are derived from genomic sequences expressed in the human
CC   bone marrow. They can be used to measure gene expression in bone marrow
CC   samples, which may enable the improved diagnosis and treatment of cancers
CC   such as lymphoma, leukaemia and myeloma. The present sequence is one of
CC   the probes of the invention.
XX
SQ   Sequence 425 BP; 112 A; 107 C; 94 G; 112 T; 0 other;

  Query Match             100.0%;  Score 21;  DB 22;  Length 425;
  Best Local Similarity   100.0%;  Pred. No. 0.69;
  Matches   21;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 AAGCCTGATTGATGACCAGGG 21
              |||||||||||||||||||||
Db        150 AAGCCTGATTGATGACCAGGG 130


RESULT 7
AAI14882/c
ID   AAI14882 standard; DNA; 425 BP.
XX
AC   AAI14882;
XX
DT   12-OCT-2001  (first entry)
XX
DE   Probe #4815 for gene expression analysis in human cervical cell sample.
XX
KW   Probe; human; microarray; gene expression; cervical epithelial cell;
KW   cervical cancer; ss.
XX
OS   Homo sapiens.
XX
PN   WO200157278-A2.
XX
PD   09-AUG-2001.
XX
PF   30-JAN-2001; 2001WO-US00670.
XX
PR   04-FEB-2000; 2000US-0180312.
PR   26-MAY-2000; 2000US-0207456.
PR   30-JUN-2000; 2000US-0608408.
PR   03-AUG-2000; 2000US-0632366.
PR   21-SEP-2000; 2000US-0234687.
PR   27-SEP-2000; 2000US-0236359.
PR   04-OCT-2000; 2000GB-0024263.
XX
PA   (MOLE-) MOLECULAR DYNAMICS INC.
XX
PI   Penn SG,  Hanzel DK,  Chen W,  Rank DR;
XX
DR   WPI; 2001-488901/53.
XX
PT   Human genome-derived single exon nucleic acid probes useful for
PT   analyzing gene expression in human cervical epithelial cells -
XX
PS   Claim 25; SEQ ID No 4815; 487pp; English.
XX
CC   The present invention relates to human single exon nucleic acid probes
CC   (SENP). The present sequence is one such probe. The SENPs are derived
CC   from human HeLa cells. The SENPs can be used to produce a single exon
CC   microarray, which can be used for measuring human gene expression in a
CC   sample derived from human cervical epithelial cells. By measuring gene
CC   expression, the probes are therefore useful in grading and/or staging
CC   of diseases of the cervix, notably cervical cancer.
CC   Note: The sequence data for this patent did not form part of the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence 425 BP; 112 A; 107 C; 94 G; 112 T; 0 other;

  Query Match             100.0%;  Score 21;  DB 22;  Length 425;
  Best Local Similarity   100.0%;  Pred. No. 0.69;
  Matches   21;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 AAGCCTGATTGATGACCAGGG 21
              |||||||||||||||||||||
Db        150 AAGCCTGATTGATGACCAGGG 130


RESULT 8
AAI36237/c
ID   AAI36237 standard; DNA; 425 BP.
XX
AC   AAI36237;
XX
DT   17-OCT-2001  (first entry)
XX
DE   Probe #4923 used to measure gene expression in human placenta sample.
XX
KW   Probe; microarray; human; placenta; antenatal diagnosis;
KW   genetic disorder; ss.
XX
OS   Homo sapiens.
XX
PN   WO200157272-A2.
XX
PD   09-AUG-2001.
XX
PF   30-JAN-2001; 2001WO-US00663.
XX
PR   04-FEB-2000; 2000US-0180312.
```

PR 26-MAY-2000; 2000US-0207456.
PR 30-JUN-2000; 2000US-0608408.
PR 03-AUG-2000; 2000US-0632366.
PR 21-SEP-2000; 2000US-0234687.
PR 27-SEP-2000; 2000US-0236359.
PR 04-OCT-2000; 2000GB-0024263.
XX
PA (MOLE-) MOLECULAR DYNAMICS INC.
XX
PI Penn SG, Hanzel DK, Chen W, Rank DR;
XX
DR WPI; 2001-488897/53.
XX
PT Human genome-derived single exon nucleic acid probes useful for
PT analyzing gene expression in human placenta -
XX
PS Claim 25; SEQ ID No 4923; 654pp; English.
XX
CC The present invention relates to single exon nucleic acid probes (SENP).
CC The present sequence is one such probe. The probes are useful for
CC producing a microarray for predicting, measuring and displaying gene
CC expression in samples derived from human placenta. The probes are useful
CC for antenatal diagnosis of human genetic disorders.
XX
SQ Sequence 425 BP; 112 A; 107 C; 94 G; 112 T; 0 other;

Query Match 100.0%; Score 21; DB 22; Length 425;
Best Local Similarity 100.0%; Pred. No. 0.69;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 AAGCCTGATTGATGACCAGGG 21
||||||||||||||||||||
Db 150 AAGCCTGATTGATGACCAGGG 130

RESULT 9
AAI04666/c
ID AAI04666 standard; DNA; 425 BP.
XX
AC AAI04666;
XX
DT 09-OCT-2001 (first entry)
XX
DE Probe #4657 used to measure gene expression in human breast sample.
XX
KW Probe; human; breast disease; breast cancer; development disorder; ss;
KW inflammatory disease; proliferative breast disease; non-carcinoma tumour.
XX
OS Homo sapiens.
XX
PN WO200157270-A2.
XX
PD 09-AUG-2001.
XX
PF 29-JAN-2001; 2001WO-US00661.
XX
PR 04-FEB-2000; 2000US-0180312.
PR 26-MAY-2000; 2000US-0207456.
PR 30-JUN-2000; 2000US-0608408.
PR 03-AUG-2000; 2000US-0632366.
PR 21-SEP-2000; 2000US-0234687.
PR 27-SEP-2000; 2000US-0236359.
PR 04-OCT-2000; 2000GB-0024263.
XX
PA (MOLE-) MOLECULAR DYNAMICS INC.
XX
PI Penn SG, Hanzel DK, Chen W, Rank DR;
XX
DR WPI; 2001-476286/51.
XX
PT Novel single exon nucleic acid probe used to measuring gene expression
PT in a human breast -
XX
PS Claim 25; SEQ ID No 4657; 322pp; English.
XX
CC The present invention relates to novel single exon nucleic acid probes.
CC The present sequence is one such probe. The probes are useful for
CC measuring human gene expression in a human breast sample, where the probe
CC hybridises at high stringency to a nucleic acid expressed in the human
CC breast. The probes are useful for predicting, diagnosing, grading,
CC staging, monitoring and prognosing diseases of the human breast,
CC particularly those diseases with polygenic aetiology. The diseases
CC include: breast cancer, disorders of development, inflammatory diseases
CC of the breast, fibrocystic changes, proliferative breast disease and
CC non-carcinoma tumours.
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 425 BP; 112 A; 107 C; 94 G; 112 T; 0 other;

Query Match 100.0%; Score 21; DB 22; Length 425;
Best Local Similarity 100.0%; Pred. No. 0.69;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 AAGCCTGATTGATGACCAGGG 21
||||||||||||||||||||
Db 150 AAGCCTGATTGATGACCAGGG 130

RESULT 10
ABS29927/c
ID ABS29927 standard; DNA; 425 BP.
XX
AC ABS29927;
XX
DT 25-FEB-2003 (first entry)
XX
DE Human liver single exon probe, SEQ ID No 4917.
XX
KW Human; single exon nucleic acid probe; liver; cirrhosis;
KW hyperlipoproteinaemia; hyperlipidaemia; hypercholesterolaemia;
KW coronary heart disease; ss.
XX
OS Homo sapiens.
XX
PN WO200157273-A2.
XX
PD 09-AUG-2001.
XX
PF 30-JAN-2001; 2001WO-US00664.
XX
PR 04-FEB-2000; 2000US-0180312.
PR 26-MAY-2000; 2000US-0207456.
PR 30-JUN-2000; 2000US-0608408.
PR 03-AUG-2000; 2000US-0632366.
PR 21-SEP-2000; 2000US-0234687.
PR 27-SEP-2000; 2000US-0236359.
PR 04-OCT-2000; 2000GB-0024263.
XX
PA (MOLE-) MOLECULAR DYNAMICS INC.
XX
PI Penn SG, Hanzel DK, Chen W, Rank DR;
XX
DR WPI; 2001-488898/53.
XX
PT Human genome-derived single exon nucleic acid probes useful for
PT analysing gene expression in human adult liver -
XX
PS Claim 1; SEQ ID No 4917; 658pp; English.
XX
CC The invention relates to a single exon nucleic acid probe (SENP) (I) for
CC measuring human gene expression in a sample derived from human adult
CC liver, comprising one of 13109 defined nucleotide sequences given in the
CC specification (or complements/ fragments). The probe hybridises at high
CC stringency to a nucleic acid molecule expressed in the human adult

```
CC   liver. (I) may be used for predicting, measuring and displaying gene
CC   expression in samples derived from human adult liver. The genes
CC   identified may be involved in genetic liver diseases such as cirrhosis,
CC   hyperlipoproteinaemia, hyperlipidaemia and hypercholesterolaemia which
CC   is associated with coronary heart disease. ABS25011-ABS51005 represent
CC   human liver single exon nucleic acid probes of the invention.
CC   Note: The sequence information for this patent does not appear in the
CC   printed specification but was obtained in electronic format directly
CC   from WIPO at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence 425 BP; 112 A; 107 C; 94 G; 112 T; 0 other;

  Query Match             100.0%;  Score 21;  DB 23;  Length 425;
  Best Local Similarity   100.0%;  Pred. No. 0.69;
  Matches   21;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 AAGCCTGATTGATGACCAGGG 21
              |||||||||||||||||||||
Db        150 AAGCCTGATTGATGACCAGGG 130


RESULT 11
ABS04866/c
ID   ABS04866 standard; DNA; 425 BP.
XX
AC   ABS04866;
XX
DT   19-AUG-2002  (first entry)
XX
DE   Human genome-derived single exon probe from lung SEQ ID No 4857.
XX
KW   Human; ds; single exon probe; asthma; lung cancer; COPD; ILD;
KW   chronic obstructive pulmonary disease; interstitial lung disease;
KW   familial idiopathic pulmonary fibrosis; neurofibromatosis;
KW   tuberous sclerosis; Gaucher's disease; Niemann-Pick disease;
KW   Hermansky-Pudlak syndrome; sarcoidosis; pulmonary haemosiderosis;
KW   pulmonary histiocytosis; lymphangioleiomyomtosis; Karagener syndrome;
KW   pulmonary alveolar proteinosis; fibrocystic pulmonary dysplasia;
KW   primary ciliary dyskinesis; pulmonary hypertension;
KW   hyaline membrane disease.
XX
OS   Homo sapiens.
XX
PN   WO200186003-A2.
XX
PD   15-NOV-2001.
XX
PF   30-JAN-2001; 2001WO-US00665.
XX
PR   04-FEB-2000; 2000US-180312P.
PR   26-MAY-2000; 2000US-207456P.
PR   30-JUN-2000; 2000US-0608408.
PR   03-AUG-2000; 2000US-0632366.
PR   21-SEP-2000; 2000US-234687P.
PR   27-SEP-2000; 2000US-236359P.
PR   04-OCT-2000; 2000GB-0024263.
XX
PA   (MOLE-) MOLECULAR DYNAMICS INC.
XX
PI   Penn SG,  Hanzel DK,  Chen W,  Rank DR;
XX
DR   WPI; 2002-114183/15.
XX
PT   Spatially-addressable set of single exon nucleic acid probes, used to
PT   measure gene expression in human lung samples -
XX
PS   Claim 1; SEQ ID No 4857; 634pp; English.
XX
CC   The invention relates to a spatially-addressable set of single exon
CC   nucleic acid probes for measuring gene expression in a sample derived
CC   from human lung comprising single exon nucleic acid probes having one of
CC   12614 nucleic acid sequences mentioned in the specification, or their
CC   complements or the 12387 open reading frames derived from the 12614
CC   probes. Also included are a microarray comprising the novel set of
CC   probes; the novel set of probes which hybridise at high stringency to a
CC   nucleic acid expressed in the human lung; measuring gene expression in a
CC   sample derived from human lung, comprising (a) contacting the array with
CC   a collection of detectably labeled nucleic acids derived from human lung
CC   mRNA, and (b) measuring the label detectably bound to each probe of
CC   the array; identifying exons in a eukaryotic genome, comprising
CC   (a) algorithmically predicting at least one exon from genomic sequences
CC   of the eukaryote; and (b) detecting specific hybridisation of detectably
CC   labeled nucleic acids from eukaryote lung mRNA, to a single exon probe,
CC   having a fragment identical to the predicted exon, the probe is included
CC   in the above mentioned microarray; assigning exons to a single gene,
CC   comprising (a) identifying exons from genomic sequence by the method
CC   above and (b) measuring the expression of each of the exons in several
CC   tissues and/or cell types using hybridisation to a single exon
CC   microarrays having a probe with the exon, where a common pattern of
CC   expression of the exons in the tissues and/or cell types indicates that
CC   the exons should be assigned to a single gene; a peptide comprising one
CC   of 12011 sequences, mentioned in the specification, or encoded by the
CC   probes/open reading frames (ORF). The probes are used for gene
CC   expression analysis, and for identifying exons in a gene, particularly
CC   using human lung derived mRNA and for the study of lung diseases
CC   such as asthma, lung cancer, chronic obstructive pulmonary disease
CC   (COPD), interstitial lung disease (ILD), familial idiopathic pulmonary
CC   fibrosis, neurofibromatosis, tuberous sclerosis, Gaucher's disease,
CC   Niemann-Pick disease, Hermansky-Pudlak syndrome, sarcoidosis, pulmonary
CC   haemosiderosis, pulmonary histiocytosis, lymphangioleiomyomtosis,
CC   pulmonary alveolar proteinosis, Karagener syndrome, fibrocystic
CC   pulmonary dysplasia, primary ciliary dyskinesis, pulmonary hypertension
CC   and hyaline membrane disease. The present sequence is a single exon
CC   probe of the invention.
CC   Note: The sequence data for this patent did not form part
CC   of the printed specification, but was obtained in electronic
CC   format directly from WIPO at
CC   ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence 425 BP; 112 A; 107 C; 94 G; 112 T; 0 other;

  Query Match             100.0%;  Score 21;  DB 24;  Length 425;
  Best Local Similarity   100.0%;  Pred. No. 0.69;
  Matches   21;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 AAGCCTGATTGATGACCAGGG 21
              |||||||||||||||||||||
Db        150 AAGCCTGATTGATGACCAGGG 130


RESULT 12
ABX95006
ID   ABX95006 standard; cDNA; 1041 BP.
XX
AC   ABX95006;
XX
DT   05-JUN-2003  (first entry)
XX
DE   cDNA encoding human tumour rejection antigen precursor, MAGE-C3.
XX
KW   TRAP; ss; tumour rejection antigen precursor; cytolytic T-cell; CTL;
KW   tumour; seminoma; bladder transitional-cell carcinoma; NSCLC; adaptor;
KW   head-and-neck squamous-cell carcinoma; breast carcinoma; sarcoma;
KW   cutaneous melanoma; nonsmall cell lung cancer; gene; MAGE-C3; human;
KW   chromosome Xq27.1-Xq27.3.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   CDS             1..1041
FT                   /*tag=  a
FT                   /product= "MAGE-C3"
XX
PN   US2002176865-A1.
XX
PD   28-NOV-2002.
```

XX
PF 01-MAR-2002; 2002US-0085108.
XX
PR 09-FEB-2000; 2000US-0501104.
PR 25-APR-1997; 97US-0845528.
PR 24-APR-1998; 98US-0066281.
PR 17-DEC-1999; 99US-0468433.
XX
PA (LUCA/) LUCAS S.
PA (BOON/) BOON-FALLEUR T.
XX
PI Lucas S, Boon-Falleur T;
XX
DR WPI; 2003-328468/31.
DR P-PSDB; ABU08932.
XX
PT Novel isolated nucleic acid encoding tumor rejection antigen precursor
PT MAGE-C3, MAGE-B5, or MAGE-B6, useful as diagnostic probes to determine
PT presence of abnormal e.g., tumor cells expressing MAGE-C1, MAGE-B5 or
PT MAGE-B6 -
XX
PS Claim 1; Fig 5; 59pp; English.
XX
CC The invention relates to an isolated nucleic acid molecule which encodes
CC a tumour rejection antigen precursor (TRAP) having an amino acid sequence
CC of a TRAP encoded by a fully defined MAGE-C3, MAGE-B5, or MAGE-B6
CC polynucleotide sequence. Also disclosed is a method which is useful for
CC determining presence of cytolytic T-cells specific for complexes of human
CC leukocyte antigen (HLA) and a peptide derived from the nucleic acid in a
CC cytotoxic T-lymphocyte (CTL)-containing sample. The nucleic acid is
CC useful as a diagnostic probe to determine the presence of abnormal
CC (tumour) cells such as seminoma, bladder transitional-cell carcinoma,
CC head-and-neck squamous-cell carcinoma, breast carcinoma, sarcoma,
CC cutaneous melanoma or nonsmall cell lung cancer (NSCLC) which express
CC MAGE-C1, MAGE-B5 or MAGE-B6. The nucleic acid is useful for diagnosing a
CC disorder characterised by expression of MAGE-C1, MAGE-B5 or MAGE-B6 TRAPs
CC or tumour rejection antigens (TRAs). The present sequence represents the
CC cDNA of the gene encoding the human tumour rejection antigen precursor,
CC MAGE-C3, which is located on chromosome Xq27.1-Xq27.3.
XX
SQ Sequence 1041 BP; 242 A; 283 C; 242 G; 274 T; 0 other;

Query Match 100.0%; Score 21; DB 25; Length 1041;
Best Local Similarity 100.0%; Pred. No. 0.81;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 AAGCCTGATTGATGACCAGGG 21
||||||||||||||||||||
Db 711 AAGCCTGATTGATGACCAGGG 731

RESULT 13
AAS88354
ID AAS88354 standard; cDNA; 7806 BP.
XX
AC AAS88354;
XX
DT 13-FEB-2002 (first entry)
XX
DE DNA encoding novel human diagnostic protein #24158.
XX
KW Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW food supplement; medical imaging; diagnostic; genetic disorder; ss.
XX
OS Homo sapiens.
XX
PN WO200175067-A2.
XX
PD 11-OCT-2001.
XX
PF 30-MAR-2001; 2001WO-US08631.
XX
PR 31-MAR-2000; 2000US-0540217.
PR 23-AUG-2000; 2000US-0649167.
XX
PA (HYSE-) HYSEQ INC.
XX
PI Drmanac RT, Liu C, Tang YT;
XX
DR WPI; 2001-639362/73.
DR P-PSDB; ABG24167.
XX
PT New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity -
XX
PS Claim 1; SEQ ID No 24158; 103pp; English.
XX
CC The invention relates to isolated polynucleotide (I) and
CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC and gene mapping, and in recombinant production of (II). The
CC polynucleotides are also used in diagnostics as expressed sequence tags
CC for identifying expressed genes. (I) is useful in gene therapy techniques
CC to restore normal activity of (II) or to treat disease states involving
CC (II). (II) is useful for generating antibodies against it, detecting or
CC quantitating a polypeptide in tissue, as molecular weight markers and as
CC a food supplement. (II) and its binding partners are useful in medical
CC imaging of sites expressing (II). (I) and (II) are useful for treating
CC disorders involving aberrant protein expression or biological activity.
CC The polypeptide and polynucleotide sequences have applications in
CC diagnostics, forensics, gene mapping, identification of mutations
CC responsible for genetic disorders or other traits to assess biodiversity
CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. AAS64197-AAS94564 represent novel human
CC diagnostic coding sequences of the invention.
CC Note: The sequence data for this patent did not appear in the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 7806 BP; 1755 A; 2185 C; 1760 G; 2106 T; 0 other;

Query Match 100.0%; Score 21; DB 23; Length 7806;
Best Local Similarity 100.0%; Pred. No. 1.2;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 AAGCCTGATTGATGACCAGGG 21
||||||||||||||||||||
Db 3971 AAGCCTGATTGATGACCAGGG 3991

RESULT 14
AAT24776
ID AAT24776 standard; cDNA to mRNA; 336 BP.
XX
AC AAT24776;
XX
DT 09-OCT-1996 (first entry)
XX
DE Human gene signature HUMGS06854.
XX
KW Gene signature; messenger RNA; mRNA; relative abundance; frequency;
KW human; cloning; mapping; non-biased library; diagnosis; detection;
KW cell typing; abnormal cell function; ss.
XX
OS Homo sapiens.
XX
PN WO9514772-A1.
XX
PD 01-JUN-1995.
XX
PF 11-NOV-1994; 94WO-JP01916.
XX
PR 12-NOV-1993; 93JP-0355504.
XX

PA (MATS/) MATSUBARA K.
PA (OKUB/) OKUBO K.
XX
PI Matsubara K, Okubo K;
XX
DR WPI; 1995-206931/27.
XX
PT Identifying gene signatures in 3'-directed human cDNA library - e.g.
PT for diagnosis of abnormal cell function, by preparing cDNA that
PT reflects relative abundance of corresp. mRNA in specific human
PT tissues
XX
PS Claim 1; Page 1692; 2245pp; Japanese.
XX
CC A single-stranded DNA (or its complementary strand or the corresp.
CC double-stranded DNA) which comprises one of the 7837 "GS" sequences
CC given in AAT19001-T26837 and which is able to hybridise to part of
CC human genomic DNA, cDNA or mRNA is claimed. The GS (Gene Signature)
CC sequences were obtained from 3'-directed cDNA libraries prepared
CC from various human tissues; synthesis of cDNA was initiated from the
CC 3'-end of mRNA by using poly(T) as the sole primer. Since the 3'-
CC untranslated sequence is unique to a particular mRNA species, almost
CC all the 3'-oriented cDNAs hybridise with specific mRNAs. Each library
CC is constructed so as to reflect accurately the relative abundance of
CC different mRNAs in the particular tissue from which it was derived.
CC The appearance frequency of a given GS in a cDNA library can be
CC determined (esp. using primers and probes derived from the GS
CC sequences) as a means of diagnosing abnormal cell function or for
CC recognising different cell types.
XX
SQ Sequence 336 BP; 91 A; 67 C; 80 G; 83 T; 15 other;

Query Match 81.0%; Score 17; DB 16; Length 336;
Best Local Similarity 100.0%; Pred. No. 73;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 5 CTGATTGATGACCAGGG 21
||||||||||||||||
Db 231 CTGATTGATGACCAGGG 247

RESULT 15
ABS69805
ID ABS69805 standard; DNA; 504 BP.
XX
AC ABS69805;
XX
DT 21-NOV-2002 (first entry)
XX
DE Novel murine polynucleotide isolated using gene trap technology #868.
XX
KW Mouse; gene trapped sequence; GTS; functional genomic analysis;
KW phage display system; gene chip; temporal gene expression;
KW tissue specific gene expression; antisense inhibition; gene targeting;
KW development disorder; cell differentiation disorder; aging; cancer;
KW autoimmune disease; lupus; inflammatory disorder; skin disorder;
KW degenerative disorder; ds.
XX
OS Mus musculus.
XX
PN US2002102543-A1.
XX
PD 01-AUG-2002.
XX
PF 30-NOV-2000; 2000US-0728445.
XX
PR 01-DEC-1999; 99US-168358P.
XX
PA (FRIE/) FRIEDRICH G.
PA (ZAMB/) ZAMBROWICZ B.
PA (SAND/) SANDS A T.
XX
PI Friedrich G, Zambrowicz B, Sands AT;
XX
DR WPI; 2002-690598/74.
XX
PT Novel murine polynucleotides that individually identify novel genes
PT into which a retroviral gene trap vector has integrated, useful in
PT genomic analysis and in discovery, development of therapeutic and
PT diagnostic agents -
XX
PS Claim 1; Page 286; 296pp; English.
XX
CC The invention describes an isolated murine polynucleotide (I) comprising
CC a contiguous stretch of at least 60 nucleotides of one of 265-677
CC nucleotide 891 OMNIBANK gene trapped sequences (GTSs) (S), given in the
CC specification. The novel genes and cells are useful in functional
CC genomic analysis and in the discovery and development of new therapeutic
CC and diagnostic agents and methods. (I) is useful for identifying the
CC coding regions of the murine genome, to isolate cDNAs, genomic clones,
CC or full-length genes/polynucleotides or homologues, heterologues,
CC paralogues, or orthologues that are capable of hybridising to one or more
CC of the GTSs under stringent conditions. (I) can be incorporated into a
CC phage display system that can be used to screen for proteins, or other
CC ligands, that are capable of binding an amino acid sequence encoded by
CC an oligonucleotide or polynucleotide sequence in at least one of the TS
CC sequences. (I) is useful in addressable arrays, such as gene chips, to
CC identify and characterise temporal and tissue specific gene expression,
CC to identify the gene of interest from many sources and for genetic
CC manipulations such as antisense inhibition and gene targeting. Decreasing
CC the level of expression of (I) and/or down regulating the activity of
CC peptides or proteins encoded by (I) is useful for treating development
CC and cell differentiation disorders, aging, cancer, autoimmune disease,
CC lupus, inflammatory disorders, skin disorders and degenerative
CC disorders. This sequence represents a murine cDNA isolated using gene
CC trap technology.
XX
SQ Sequence 504 BP; 110 A; 152 C; 146 G; 96 T; 0 other;

Query Match 80.0%; Score 16.8; DB 24; Length 504;
Best Local Similarity 90.0%; Pred. No. 1e+02;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 2 AGCCTGATTGATGACCAGGG 21
||| ||||||||||||||
Db 461 AGCAGGATTGATGACCAGGG 480

Search completed: February 19, 2004, 21:28:33
Job time : 176.5 secs

GenCore version 5.1.6

OM nucleic - nucleic search, using sw model

Run on: February 19, 2004, 20:01:08 ; Search time 1596.5 Seconds
(without alignments)
538.116 Million cell updates/sec

Title: US-10-085-108-21_COPY_711_731
Perfect score: 21
Sequence: 1 AAGCCTGATTGATGACCAGGG 21

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 2888711 seqs, 20454813386 residues

Total number of hits satisfying chosen parameters: 5777422

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :
GenEmbl:*
1: gb_ba:*
2: gb_htg:*
3: gb_in:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_sts:*
12: gb_sy:*
13: gb_un:*
14: gb_vi:*
15: em_ba:*
16: em_fun:*
17: em_hum:*
18: em_in:*
19: em_mu:*
20: em_om:*
21: em_or:*
22: em_ov:*
23: em_pat:*
24: em_ph:*
25: em_pl:*
26: em_ro:*
27: em_sts:*
28: em_un:*
29: em_vi:*
30: em_htg_hum:*
31: em_htg_inv:*
32: em_htg_other:*
33: em_htg_mus:*
34: em_htg_pln:*
35: em_htg_rod:*
36: em_htg_mam:*
37: em_htg_vrt:*
38: em_sy:*
39: em_htgo_hum:*
40: em_htgo_mus:*
41: em_htgo_other:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.		Score	% Query Match	Length	DB	ID	Description
1		21	100.0	1932	9	AF490508	AF490508 Homo sapi
2		21	100.0	124497	9	HS232G24	AL022152 Human DNA
3	c	19	90.5	153439	2	BX323047	BX323047 Danio rer
4	c	19	90.5	159948	2	BX248240	BX248240 Danio rer
5		17.8	84.8	1494	9	AK098443	AK098443 Homo sapi
6		17.8	84.8	1681	9	BC048268	BC048268 Homo sapi
7	c	17.8	84.8	67201	9	AL136169	AL136169 Human DNA
8		17.8	84.8	163746	2	AC102710	AC102710 Mus muscu
9	c	17.8	84.8	174191	9	AC109462	AC109462 Homo sapi
10	c	17.8	84.8	177733	2	AC113908	AC113908 Rattus no
11	c	17.8	84.8	180353	2	AC128288	AC128288 Rattus no
12	c	17.8	84.8	186118	2	AC117856	AC117856 Rattus no
13		17.8	84.8	194454	2	AC111040	AC111040 Mus muscu
14	c	17.8	84.8	194520	10	AL691424	AL691424 Mouse DNA
15		17.8	84.8	194636	10	AC123058	AC123058 Mus muscu
16	c	17.8	84.8	214493	2	AC097047	AC097047 Rattus no
17	c	17.8	84.8	227369	2	AC114469	AC114469 Rattus no
18	c	17.8	84.8	231827	2	AC096130	AC096130 Rattus no
19		17.8	84.8	232476	2	AC118621	AC118621 Mus muscu
20	c	17.8	84.8	243257	2	AC130512	AC130512 Rattus no
21		17.8	84.8	251759	2	AC137283	AC137283 Rattus no
22		17.8	84.8	257104	2	AC122547	AC122547 Mus muscu
23	c	17.8	84.8	274024	2	AC102995	AC102995 Rattus no
24	c	17.8	84.8	274467	2	AC113707	AC113707 Rattus no
25	c	17.4	82.9	196	5	AY185781	AY185781 Ambystoma
26	c	17.4	82.9	197	5	AY185782	AY185782 Ambystoma
27	c	17.4	82.9	223	5	AY185785	AY185785 Ambystoma
28	c	17.4	82.9	223	5	AY185786	AY185786 Ambystoma
29	c	17.4	82.9	224	5	AY185783	AY185783 Ambystoma
30	c	17.4	82.9	224	5	AY185787	AY185787 Ambystoma
31	c	17.4	82.9	227	5	AY185784	AY185784 Ambystoma
32	c	17.4	82.9	4202	5	AF001958	AF001958 Ambystoma
33		17.4	82.9	92337	10	AL604022	AL604022 Mouse DNA
34	c	17.4	82.9	120492	3	CEY43F8C	AL032637 Caenorhab
35		17.4	82.9	198727	2	AC122401	AC122401 Mus muscu
36	c	17.4	82.9	203961	9	AC011374	AC011374 Homo sapi
37	c	17.4	82.9	218092	2	AC019272	AC019272 Mus muscu
38	c	17.4	82.9	219491	2	AC137244	AC137244 Rattus no
39		17.4	82.9	226771	2	AC110940	AC110940 Rattus no
40		17.4	82.9	242602	2	AC095427	AC095427 Rattus no
41		17	81.0	108542	10	AL627307	AL627307 Mouse DNA
42		17	81.0	208966	2	AC099588	AC099588 Mus muscu
43	c	17	81.0	212119	2	AC099602	AC099602 Mus muscu
44		17	81.0	236604	2	AC099619	AC099619 Mus muscu
45	c	16.8	80.0	403	9	AF490961	AF490961 Homo sapi

ALIGNMENTS

RESULT 1
AF490508
LOCUS AF490508 1932 bp mRNA linear PRI 04-APR-2002
DEFINITION Homo sapiens hepatocellular carcinoma-associated protein HCA2 mRNA, complete cds.
ACCESSION AF490508
VERSION AF490508.1 GI:19919741
KEYWORDS .
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 1932)
AUTHORS Dong,X. and Chen,W.
TITLE Identification of genes in the chromosome X that are differentially

```
            expressed in hepatocellular carcinoma
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 1932)
  AUTHORS   Dong,X. and Chen,W.
  TITLE     Direct Submission
  JOURNAL   Submitted (06-MAR-2002) Department of Immunology, T Cell Research
            Lab., Xueuan Road 38, Beijing 100083, China
FEATURES             Location/Qualifiers
     source          1. .1932
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /chromosome="X"
     CDS             1. .1932
                     /codon_start=1
                     /product="hepatocellular carcinoma-associated protein
                     HCA2"
                     /protein_id="AAM08355.1"
                     /db_xref="GI:19919742"
                     /translation="MLLPCHWVLDATFSDGSLGQWVKNTCATYALSPVVLPPQPQPRK
                     KATDKDYSAFHLGHLREVRLFLRGGTSDQRMDSLVLCPTYFKLWRTLSGSPGLQLSDL
                     HFGSQPEGKFSLRRAVSVKQREEPQDWPLNEKRTLWKDSDLPTWRRGTGYTLSLPAVS
                     PGKRLWGEKAGSLPESEPLFTYTLDEKVDKLVQFLLLKYQAKEPLTRAEMQMNVINTY
                     TGYFPMIFRKAREFIEILFGISLTEVDPDHFYVFVNTLDLTCEGSLSDEQGMPQNRLL
                     ILILSVIFIKGNCASEEVIWEVLNAIGPWSALAGFADVLSRLALWESEGPEAFCEESG
                     LRSAEGSVLDLANPQGLAGHRQEDGRRGLTEASPQQKKGGEDEDMPAAGMPPLPQSPP
                     EIPPQGPPKISPQGPPQSPPQSPLDSCSSPLLWTRLDEESSSEEEDTATWHALPESES
                     LPRYALDEKVAELVQFLLLKYQTKEPVTKAEMLTTVIKKYKDYFPMIFGKAHEFIELI
                     FGIALTDMDPDNHSYFFEDTLDLTYEGSLIDDQGMPKNCLLILILSMIFIKGSCVPEE
                     VIWEVLSAIGPIQRPAREVLEFLSKLSSIIPSAFPSWYMDALKDMEDRAQAIIDTTDD
                     ATAMASASPSVMSTNFCPE"
BASE COUNT      477 a    489 c    510 g    456 t
ORIGIN

  Query Match             100.0%;  Score 21;  DB 9;  Length 1932;
  Best Local Similarity   100.0%;  Pred. No. 13;
  Matches   21;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 AAGCCTGATTGATGACCAGGG 21
              |||||||||||||||||||||
Db       1605 AAGCCTGATTGATGACCAGGG 1625


RESULT 2
HS232G24
LOCUS       HS232G24              124497 bp    DNA     linear   PRI 09-MAR-2002
DEFINITION  Human DNA sequence from clone RP6-232G24 on chromosome Xq27.1-27.3
            Contains the gene for tha melanoma antigen gene family protein,
            MAGEC3 and the MAGEC1 gene for melanoma antigen, family C,1,
            complete sequence.
ACCESSION   AL022152
VERSION     AL022152.1  GI:3150086
KEYWORDS    HTG; MAGEC1; MAGEC3.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 124497)
  AUTHORS   Bird,C.
  TITLE     Direct Submission
  JOURNAL   Submitted (08-MAR-2002) Wellcome Trust Sanger Institute, Hinxton,
            Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
            humquery@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
COMMENT     On May 22, 1998 this sequence version replaced gi:2969932.
            During sequence assembly data is compared from overlapping clones.
            Where differences are found these are annotated as variations
            together with a note of the overlapping clone name. Note that the
            variation annotation may not be found in the sequence submission
            corresponding to the overlapping clone, as we submit sequences with
            only a small overlap as described above.
            This sequence was finished as follows unless otherwise noted: all
            regions were either double-stranded or sequenced with an alternate
            chemistry or covered by high quality data (i.e., phred quality >=
            30); an attempt was made to resolve all sequencing problems, such
            as compressions and repeats; all regions were covered by at least
            one plasmid subclone or more than one M13 subclone; and the
            assembly was confirmed by restriction digest. The following
            abbreviations are used to associate primary accession numbers given
            in the feature table with their source databases: Em:, EMBL; Sw:,
            SWISSPROT; Tr:, TREMBL; Wp:, WORMPEP; Information on the WORMPEP
            database can be found at
            http://www.sanger.ac.uk/Projects/C_elegans/wormpep This sequence
            was generated from part of bacterial clone contigs of human
            chromosome X, constructed by the Sanger Centre Chromosome X Mapping
            Group.  Further information can be found at
            http://www.sanger.ac.uk/HGP/ChrX
            RP6-232G24 is from the library RPCI-6 constructed by the group of
            Pieter de Jong. For further details see
            http://www.chori.org/bacpac/home.htm
            VECTOR: pPAC4
            This sequence is the entire insert of clone RP6-232G24 The true
            right end of clone RP3-326L12 is at 82755 in this sequence.
FEATURES             Location/Qualifiers
     source          1. .124497
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="RZPD:RPCIP709G24232"
                     /db_xref="taxon:9606"
                     /chromosome="X"
                     /map="q27.1-27.3"
                     /clone="RP6-232G24"
                     /clone_lib="RPCI-6"
     repeat_region   6. .354
                     /note="L1MC2 repeat: matches 5195. .5546 of consensus"
     repeat_region   356. .762
                     /note="L1MC2 repeat: matches 5843. .6246 of consensus"
     repeat_region   763. .794
                     /note="16 copies 2 mer tg 93% conserved"
     repeat_region   946. .1161
                     /note="L1ME1 repeat: matches 5354. .5569 of consensus"
     repeat_region   1181. .1399
                     /note="L1ME1 repeat: matches 5914. .6136 of consensus"
     repeat_region   1471. .1606
                     /note="L2 repeat: matches 2574. .2708 of consensus"
     repeat_region   1939. .2330
                     /note="14 copies 28 mer 57% conserved"
     repeat_region   1969. .2326
                     /note="179 copies 2 mer gg 57% conserved"
     repeat_region   1972. .2331
                     /note="20 copies 18 mer 59% conserved"
     repeat_region   1996. .2265
                     /note="5 copies 54 mer 65% conserved"
     repeat_region   2029. .2181
                     /note="3 copies 51 mer 74% conserved"
     repeat_region   2153. .2347
                     /note="5 copies 39 mer 65% conserved"
     gene            2426. .4801
                     /gene="dA232G24.1"
     mRNA            join(2426. .2597,3291. .3384,3693. .4801)
                     /gene="dA232G24.1"
                     /product="dA232G24.1 (Melanoma antigen gene family
                     protein, MAGEC3)"
                     /note="(possible pseudogene)
                     match: cDNAs: Em:AF151378
                     match: ESTs: Em:BI460078"
                     /evidence=not_experimental
     repeat_region   2827. .2872
                     /note="23 copies 2 mer gg 78% conserved"
     CDS             3761. .4801
                     /gene="dA232G24.1"
                     /note="(possible pseudogene)
                     match: proteins: Sw:P43363 Tr:Q9BZ80 Tr:O60781 Tr:AAH04105
                     Tr:O95529 Tr:Q9R2A2 Sw:O15479 Tr:O89006 Tr:Q9BUN9
                     Tr:Q96M61"
                     /codon_start=1
                     /evidence=not_experimental
```

```
                /product="dA232G24.1 (Melanoma antigen gene family
                protein, MAGEC3)"
                /protein_id="CAA18146.1"
                /db_xref="GI:3150087"
                /translation="MPLFPNLPRLSFEEDFQNPSVTEDLVDAQDSIDEEEEDASSTSS
                SSFHFLFPSSSSLSSSSPLSSPLPSTLILGVPEDEDMPAAGMPPLPQSPPEIPPQGPP
                KISPQGPPQSPPQSPLDSCSSPLLWTRLDEESSSEEEDTATWHALPESESLPRYALDE
                KVAELVQFLLLKYQTKEPVTKAEMLTTVIKKYKDYFPMIFGKAHEFIELIFGIALTDM
                DPDNHSYFFEDTLDLTYEGSLIDDQGMPKNCLLILILSMIFIKGSCVPEEVIWEVLSA
                IGVCAGREHFIYGDPRKLLTIHWVQRKYLEYREVPNSAPPRYEFLWGPRAHSEASKRS
                LRVFIQAIQYHP"
repeat_region   3876. .3959
                /note="7 copies 12 mer 72% conserved"
repeat_region   4027. .4110
                /note="7 copies 12 mer 88% conserved"
misc_feature    complement(5293. .5441)
                /note="match: GSS: Em:AQ060261"
repeat_region   6422. .6624
                /note="MIR repeat: matches 47. .261 of consensus"
repeat_region   8315. .8700
                /note="MLT1B repeat: matches 1. .390 of consensus"
repeat_region   9138. .9250
                /note="MLT1B repeat: matches 1. .130 of consensus"
repeat_region   9287. .9731
                /note="L1MC3 repeat: matches 7166. .7618 of consensus"
gene            11002. .16491
                /gene="MAGEC1"
mRNA            join(11002. .11086,11825. .11922,12078. .12184,12517. .16491)
                /gene="MAGEC1"
                /product="dA232G24.2 (Melanoma antigen gene family
                protein, MAGEC1)"
                /note="match: cDNAs: Em:AF056334
                match: ESTs: Em:AI652057 Em:AL121366 Em:AI126114
                Em:BG480822"
                /evidence=not_experimental
misc_feature    complement(11962. .12294)
                /note="match: GSS: Em:AQ064665"
CDS             join(12181. .12184,12517. .15941)
                /gene="MAGEC1"
                /note="match: proteins: Sw:O60732 Tr:O95529 Sw:Q9UBF1
                Tr:Q96D45 Tr:Q99NC2 Tr:Q99PH7 Sw:Q9Y5V3"
                /codon_start=1
                /evidence=not_experimental
                /product="dA232G24.2 (Melanoma antigen gene family
                protein, MAGEC1)"
                /protein_id="CAD27434.1"
                /db_xref="GI:20095260"
                /translation="MGDKDMPTAGMPSLLQSSSESPQSCPEGEDSQSPLQIPQSSPES
                DDTLYPLQSPQSRSEGEDSSDPLQRPPEGKDSQSPLQIPQSSPEGDDTQSPLQNSQSS
                PEGKDSLSPLEISQSPPEGEDVQSPLQNPASSFFSSALLSIFQSSPESTQSPFEGFPQ
                SVLQIPVSAASSSTLVSIFQSSPESTQSPFEGFPQSPLQIPVSRSFSSTLLSIFQSSP
                ERTQSTFEGFAQSPLQIPVSPSSSSTLLSLFQSFSERTQSTFEGFAQSSLQIPVSPSF
                SSTLVSLFQSSPERTQSTFEGFPQSPLQIPVSSSSSSTLLSLFQSSPERTHSTFEGFP
                QSLLQIPMTSSFSSTLLSIFQSSPESAQSTFEGFPQSPLQIPGSPSFSSTLLSLFQSS
                PERTHSTFEGFPQSPLQIPMTSSFSSTLLSILQSSPESAQSAFEGFPQSPLQIPVSSS
                FSYTLLSLFQSSPERTHSTFEGFPQSPLQIPVSSSSSSSTLLSLFQSSPECTQSTFEG
                FPQSPLQIPQSPPEGENTHSPLQIVPSLPEWEDSLSPHYFPQSPPQGEDSLSPHYFPQ
                SPPQGEDSLSPHYFPQSPQGEDSLSPHYFPQSPPQGEDSMSPLYFPQSPLQGEEFQSS
                LQSPVSICSSSTPSSLPQSFPESSQSPPEGPVQSPLHSPQSPPEGMHSQSPLQSPESA
                PEGEDSLSPLQIPQSPLEGEDSLSSLHFPQSPPEWEDSLSPLHFPQFPPQGEDFQSSL
                QSPVSICSSSTSLSLPQSFPESPQSPPEGPAQSPLQRPVSSFFSYTLASLLQSSHESP
                QSPPEGPAQSPLQSPVSSFPSSTSSSLSQSSPVSSFPSSTSSSLSKSSPESPLQSPVI
                SFSSSTSLSPFSEESSSPVDEYTSSSDTLLESDSLTDSESLIESEPLFTYTLDEKVDE
                LARFLLLKYQVKQPITKAEMLTNVISRYTGYFPVIFRKAREFIEILFGISLREVDPDD
                SYVFVNTLDLTSEGCLSDEQGMSQNRLLILILSIIFIKGTYASEEVIWDVLSGIGVRA
                GREHFAFGEPRELLTKVWVQEHYLEYREVPNSSPPRYEFLWGPRAHSEVIKRKVVEFL
                AMLKNTVPITFPSSYKDALKDVEERAQAIIDTTDDSTATESASSSVMSPSFSSE"
repeat_region   12417. .12500
                /note="7 copies 12 mer 70% conserved"
repeat_region   12420. .12497
                /note="13 copies 6 mer tcctcc 67% conserved"
repeat_region   12675. .12752
                /note="2 copies 39 mer 89% conserved"
repeat_region   12721. .12912
                /note="4 copies 48 mer 88% conserved"
repeat_region   12928. .13977
                /note="10 copies 105 mer 94% conserved"
repeat_region   13844. .14719
                /note="2 copies 438 mer 82% conserved"
repeat_region   14023. .14262
                /note="5 copies 48 mer 84% conserved"
repeat_region   14039. .14266
                /note="19 copies 12 mer 58% conserved"
repeat_region   14217. .14306
                /note="2 copies 45 mer 100% conserved"
repeat_region   14455. .14742
                /note="24 copies 12 mer 56% conserved"
repeat_region   14480. .14719
                /note="5 copies 48 mer 89% conserved"
repeat_region   14683. .14997
                /note="3 copies 105 mer 86% conserved"
repeat_region   14948. .15049
                /note="2 copies 51 mer 99% conserved"
repeat_region   15169. .15222
                /note="3 copies 18 mer 88% conserved"
repeat_region   17466. .17718
                /note="MIR repeat: matches 11. .261 of consensus"
repeat_region   19287. .19670
                /note="MLT1B repeat: matches 1. .390 of consensus"

  Query Match             100.0%;  Score 21;  DB 9;  Length 124497;
  Best Local Similarity   100.0%;  Pred. No. 11;
  Matches   21;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 AAGCCTGATTGATGACCAGGG 21
              |||||||||||||||||||||
Db       4471 AAGCCTGATTGATGACCAGGG 4491


RESULT 3
BX323047/c
LOCUS       BX323047              153439 bp    DNA     linear   HTG 08-APR-2003
DEFINITION  Danio rerio clone RP71-76P17, *** SEQUENCING IN PROGRESS ***, 13
            unordered pieces.
ACCESSION   BX323047
VERSION     BX323047.3  GI:29691743
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE      Danio rerio (zebrafish)
  ORGANISM  Danio rerio
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Ostariophysi;
            Cypriniformes; Cyprinidae; Danio.
REFERENCE   1  (bases 1 to 153439)
  AUTHORS   Burton,J.
  TITLE     Direct Submission
  JOURNAL   Submitted (07-APR-2003) Wellcome Trust Sanger Institute, Hinxton,
            Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
            zfish-help@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
COMMENT     On Apr 9, 2003 this sequence version replaced gi:29603199.
            -------------- Genome Center
            Center: Wellcome Trust Sanger Institute
            Center code: SC
            Web site: http://www.sanger.ac.uk
            Contact: zfish-help@sanger.ac.uk
            -------------- Project Information
            Center project name: bZ76P17
            -------------- Summary Statistics
            Assembly program: XGAP4; version 4.5
            Chemistry: Dye-terminator; 100% of reads
            Consensus quality: 151038 bases at least Q40
            Consensus quality: 151684 bases at least Q30
            Consensus quality: 151984 bases at least Q20
            Insert size: 152239; sum-of-contigs
            Insert size: 165247; 2.1% error; agarose-fp
            Quality coverage: 4.94x in Q20 bases; sum-of-contigs Quality
            coverage: 4.97x in Q20 bases; agarose-fp
```

```
              --------------.
              * NOTE: This is a 'working draft' sequence. It currently
              * consists of 13 contigs. The true order of the pieces
              * is not known and their order in this sequence record is
              * arbitrary. Gaps between the contigs are represented as
              * runs of N, but the exact sizes of the gaps are unknown.
              * This record will be updated with the finished sequence
              * as soon as it is available and the accession number will
              * be preserved.
              *        1     3298: contig of 3298 bp in length
              *     3299     3398: gap of 100 bp
              *     3399    10608: contig of 7210 bp in length
              *    10609    10708: gap of 100 bp
              *    10709    37823: contig of 27115 bp in length
              *    37824    37923: gap of 100 bp
              *    37924    45375: contig of 7452 bp in length
              *    45376    45475: gap of 100 bp
              *    45476    58730: contig of 13255 bp in length
              *    58731    58830: gap of 100 bp
              *    58831    61226: contig of 2396 bp in length
              *    61227    61326: gap of 100 bp
              *    61327    64481: contig of 3155 bp in length
              *    64482    64581: gap of 100 bp
              *    64582    84036: contig of 19455 bp in length
              *    84037    84136: gap of 100 bp
              *    84137    89161: contig of 5025 bp in length
              *    89162    89261: gap of 100 bp
              *    89262   113906: contig of 24645 bp in length
              *   113907   114006: gap of 100 bp
              *   114007   142019: contig of 28013 bp in length
              *   142020   142119: gap of 100 bp
              *   142120   149275: contig of 7156 bp in length
              *   149276   149375: gap of 100 bp
              *   149376   153439: contig of 4064 bp in length.
FEATURES             Location/Qualifiers
     source          1. .153439
                     /organism="Danio rerio"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:7955"
                     /clone="RP71-76P17"
                     /clone_lib="RPCI-71"
     misc_feature    1. .3298
                     /note="assembly_fragment:00374
                     fragment_chain:1"
     misc_feature    3399. .10608
                     /note="assembly_fragment:00202
                     fragment_chain:1"
     misc_feature    10709. .37823
                     /note="assembly_fragment:00149.0"
     misc_feature    37924. .45375
                     /note="assembly_fragment:00149.1"
     misc_feature    45476. .58730
                     /note="assembly_fragment:00440.0"
     misc_feature    58831. .61226
                     /note="assembly_fragment:00440.1"
     misc_feature    61327. .64481
                     /note="assembly_fragment:00440.2"
     misc_feature    64582. .84036
                     /note="assembly_fragment:00440.3"
     misc_feature    84137. .89161
                     /note="assembly_fragment:00707.0"
     misc_feature    89262. .113906
                     /note="assembly_fragment:00707.1"
     misc_feature    114007. .142019
                     /note="assembly_fragment:00707.2"
     misc_feature    142120. .149275
                     /note="assembly_fragment:00861.0"
     misc_feature    149376. .153439
                     /note="assembly_fragment:00861.1"
BASE COUNT    32783 a  43384 c  43236 g  32836 t   1200 others
ORIGIN

  Query Match             90.5%;  Score 19;  DB 2;  Length 153439;
```

```
  Best Local Similarity   100.0%;  Pred. No. 1.1e+02;
  Matches   19;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          3 GCCTGATTGATGACCAGGG 21
              |||||||||||||||||||
Db     138696 GCCTGATTGATGACCAGGG 138678


RESULT 4
BX248240/c
LOCUS       BX248240              159948 bp    DNA     linear   HTG 03-APR-2003
DEFINITION  Danio rerio clone CH211-162I19, *** SEQUENCING IN PROGRESS ***, 8
            unordered pieces.
ACCESSION   BX248240
VERSION     BX248240.3  GI:29539163
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE      Danio rerio (zebrafish)
  ORGANISM  Danio rerio
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Ostariophysi;
            Cypriniformes; Cyprinidae; Danio.
REFERENCE   1  (bases 1 to 159948)
  AUTHORS   Mclaren,S.
  TITLE     Direct Submission
  JOURNAL   Submitted (03-APR-2003) Wellcome Trust Sanger Institute, Hinxton,
            Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
            zfish-help@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
COMMENT     On Apr 3, 2003 this sequence version replaced gi:28268576.
            -------------- Genome Center
            Center: Wellcome Trust Sanger Institute
            Center code: SC
            Web site: http://www.sanger.ac.uk
            Contact: zfish-help@sanger.ac.uk
            -------------- Project Information
            Center project name: zC162I19
            -------------- Summary Statistics
            Assembly program: XGAP4; version 4.5
            Chemistry: Dye-terminator; 100% of reads
            Consensus quality: 156495 bases at least Q40
            Consensus quality: 157186 bases at least Q30
            Consensus quality: 158113 bases at least Q20
            Insert size: 159248; sum-of-contigs
            Insert size: 152025; 6.7% error; agarose-fp
            Quality coverage: 9.26x in Q20 bases; sum-of-contigs Quality
            coverage: 9.80x in Q20 bases; agarose-fp
            --------------.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 8 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1     3108: contig of 3108 bp in length
            *     3109     3208: gap of 100 bp
            *     3209    32890: contig of 29682 bp in length
            *    32891    32990: gap of 100 bp
            *    32991    49822: contig of 16832 bp in length
            *    49823    49922: gap of 100 bp
            *    49923    55063: contig of 5141 bp in length
            *    55064    55163: gap of 100 bp
            *    55164   111844: contig of 56681 bp in length
            *   111845   111944: gap of 100 bp
            *   111945   131419: contig of 19475 bp in length
            *   131420   131519: gap of 100 bp
            *   131520   149030: contig of 17511 bp in length
            *   149031   149130: gap of 100 bp
            *   149131   159948: contig of 10818 bp in length.
FEATURES             Location/Qualifiers
     source          1. .159948
                     /organism="Danio rerio"
                     /mol_type="genomic DNA"
```

```
                     /db_xref="taxon:7955"
                     /clone="CH211-162I19"
                     /clone_lib="CHORI-211"
     misc_feature    1. .3108
                     /note="assembly_fragment:01033
                     fragment_chain:1
                     clone_end:SP6
                     vector_side:left"
     misc_feature    3209. .32890
                     /note="assembly_fragment:01691
                     fragment_chain:1"
     misc_feature    32991. .49822
                     /note="assembly_fragment:00994
                     fragment_chain:1"
     misc_feature    49923. .55063
                     /note="assembly_fragment:00405
                     fragment_chain:1"
     misc_feature    55164. .111844
                     /note="assembly_fragment:01949
                     fragment_chain:1"
     misc_feature    111945. .131419
                     /note="assembly_fragment:00155
                     fragment_chain:1"
     misc_feature    131520. .149030
                     /note="assembly_fragment:00672"
     misc_feature    149131. .159948
                     /note="assembly_fragment:00468
                     clone_end:T7
                     vector_side:right"
BASE COUNT    51499 a  27454 c  27877 g  52417 t    701 others
ORIGIN

  Query Match             90.5%;  Score 19;  DB 2;  Length 159948;
  Best Local Similarity   100.0%;  Pred. No. 1.1e+02;
  Matches   19;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 AAGCCTGATTGATGACCAG 19
              ||||||||||||||||||||
Db      92696 AAGCCTGATTGATGACCAG 92678


RESULT 5
AK098443
LOCUS       AK098443               1494 bp    mRNA    linear   PRI 15-JUL-2002
DEFINITION  Homo sapiens cDNA FLJ25577 fis, clone JTH07710.
ACCESSION   AK098443
VERSION     AK098443.1  GI:21758456
KEYWORDS    oligo capping; fis (full insert sequence).
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1
  AUTHORS   Ninomiya,K., Wagatsuma,M., Kanda,K., Kondo,H., Yokoi,T.,
            Kodaira,H., Furuya,T., Takahashi,M., Kikkawa,E., Omura,Y., Abe,K.,
            Kamihara,K., Katsuta,N., Sato,K., Tanikawa,M., Yamazaki,M.,
            Suzuki,Y., Hata,H., Nakagawa,K., Mizuno,S., Morinaga,M.,
            Kawamura,M., Sugiyama,T., Irie,R., Otsuki,T., Sato,H.,
            Nishikawa,T., Sugiyama,A., Kawakami,B., Nagai,K., Isogai,T. and
            Sugano,S.
  TITLE     NEDO human cDNA sequencing project
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 1494)
  AUTHORS   Sugano,S. and Suzuki,Y.
  TITLE     Direct Submission
  JOURNAL   Submitted (08-JUL-2002) Sumio Sugano, Institute of Medical Science,
            University of Tokyo, Laboratory of Genome Structure, Human Genome
            Center; Shirokane-dai, 4-6-1, Minato-ku, Tokyo 108-8639, Japan
            (E-mail:cdna1@ims.u-tokyo.ac.jp, Tel:81-3-5449-5286,
            Fax:81-3-5449-5416)
COMMENT     NEDO human cDNA sequencing project supported by Ministry of
            Economy, Trade and Industry of Japan; cDNA full insert sequencing:
            Research Association for Biotechnology (RAB); cDNA library
            construction and 5'-end one pass sequencing: Institute of Medical
            Science, University of Tokyo, Laboratory of Genome Structure, Human
            Genome Center; 3'-end one pass sequencing: RAB; clone selection for
            full insert sequencing: RAB and Helix Research Institute.
FEATURES             Location/Qualifiers
     source          1. .1494
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="JTH07710"
                     /cell_line="JCR"
                     /tissue_type="thyroid"
                     /clone_lib="JTH"
                     /note="cloning vector: pME18SFL3"
BASE COUNT      368 a    394 c    292 g    440 t
ORIGIN

  Query Match             84.8%;  Score 17.8;  DB 9;  Length 1494;
  Best Local Similarity   90.5%;  Pred. No. 5.5e+02;
  Matches   19;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy          1 AAGCCTGATTGATGACCAGGG 21
              || |||| ||||||||||||||
Db       1244 AAACCTGGTTGATGACCAGGG 1264


RESULT 6
BC048268
LOCUS       BC048268               1681 bp    mRNA    linear   PRI 31-MAR-2003
DEFINITION  Homo sapiens, clone IMAGE:4512974, mRNA.
ACCESSION   BC048268
VERSION     BC048268.1  GI:29387246
KEYWORDS    .
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 1681)
  AUTHORS   Strausberg,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (07-MAR-2003) National Institutes of Health, Mammalian
            Gene Collection (MGC), Cancer Genomics Office, National Cancer
            Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
            USA
  REMARK    NIH-MGC Project URL: http://mgc.nci.nih.gov
COMMENT     Contact: MGC help desk
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: ATCC
            cDNA Library Preparation: Life Technologies, Inc.
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: National Institutes of Health Intramural
            Sequencing Center (NISC),
            Gaithersburg, Maryland;
            Web site: http://www.nisc.nih.gov/
            Contact: nisc_mgc@nhgri.nih.gov
            Akhter,N., Ayele,K., Beckstrom-Sternberg,S.M., Benjamin,B.,
            Blakesley,R.W., Bouffard,G.G., Breen,K., Brinkley,C., Brooks,S.,
            Dietrich,N.L., Granite,S., Guan,X., Gupta,J., Haghighi,P.,
            Hansen,N., Ho,S.-L., Karlins,E., Kwong,P., Laric,P., Legaspi,R.,
            Maduro,Q.L., Masiello,C., Maskeri,B., Mastrian,S.D.,McCloskey,J.C.,
            McDowell,J., Pearson,R., Stantripop,S., Thomas,P.J., Touchman,J.W.,
            Tsurgeon,C., Vogt,J.L., Walker,M.A., Wetherby,K.D., Wiggins,L.,
            Young,A., Zhang,L.-H. and Green,E.D.

            Clone distribution: MGC clone distribution information can be found
            through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov
            Series: IRAK Plate: 93 Row: d Column: 5
            This clone was selected for full length sequencing because it
            passed the following selection criteria: Similarity but not
            identity to protein.
FEATURES             Location/Qualifiers
     source          1. .1681
                     /organism="Homo sapiens"
```

```
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:4512974"
                     /tissue_type="Bladder, transitional cell papilloma"
                     /clone_lib="NIH_MGC_93"
                     /lab_host="DH10B"
                     /note="Vector: pCMV-SPORT6"
BASE COUNT       404 a    449 c    357 g    471 t
ORIGIN

  Query Match             84.8%;  Score 17.8;  DB 9;  Length 1681;
  Best Local Similarity   90.5%;  Pred. No. 5.5e+02;
  Matches   19;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy          1 AAGCCTGATTGATGACCAGGG 21
              || |||| ||||||||||||||
Db       1414 AAACCTGGTTGATGACCAGGG 1434


RESULT 7
AL136169/c
LOCUS       AL136169               67201 bp    DNA     linear   PRI 05-JUN-2003
DEFINITION  Human DNA sequence from clone RP4-809E13 on chromosome Xq26.1-27.1
            Contains the gene for a putative prenylated protein, two putative
            prenylated protein pseudogenes, ESTs, STSs, GSSs and three putative
            CpG islands, complete sequence.
ACCESSION   AL136169
VERSION     AL136169.6  GI:7160488
KEYWORDS    HTG; CpG island.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 67201)
  AUTHORS   Heath,P.
  TITLE     Direct Submission
  JOURNAL   Submitted (05-JUN-2003) Wellcome Trust Sanger Institute, Hinxton,
            Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
            humquery@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
COMMENT     On Mar 6, 2000 this sequence version replaced gi:6996160.
            During sequence assembly data is compared from overlapping clones.
            Where differences are found these are annotated as variations
            together with a note of the overlapping clone name. Note that the
            variation annotation may not be found in the sequence submission
            corresponding to the overlapping clone, as we submit sequences with
            only a small overlap as described above.
            The following abbreviations are used to associate primary accession
            numbers given in the feature table with their source databases:
            Em:, EMBL; Sw:, SWISSPROT; Tr:, TREMBL; Wp:, WORMPEP; Information
            on the WORMPEP database can be found at
            http://www.sanger.ac.uk/Projects/C_elegans/wormpep This sequence
            was generated from part of bacterial clone contigs of human
            chromosome X, constructed by the Sanger Centre Chromosome X Mapping
            Group.  Further information can be found at
            http://www.sanger.ac.uk/HGP/ChrX
            -------------- Genome Center
            Center: Wellcome Trust Sanger Institute
            Center code: SC
            Web site: http://www.sanger.ac.uk
            Contact: humquery@sanger.ac.uk
            --------------
            This sequence was finished as follows unless otherwise noted: all
            regions were either double-stranded or sequenced with an alternate
            chemistry or covered by high quality data (i.e., phred quality >=
            30); an attempt was made to resolve all sequencing problems, such
            as compressions and repeats; all regions were covered by at least
            one plasmid subclone or more than one M13 subclone; and the
            assembly was confirmed by restriction digest, except on the rare
            occasion of the clone being a YAC.
            RP4-809E13 is from the library RPCI-4 constructed by the group of
            Pieter de Jong. For further details see
            http://www.chori.org/bacpac/home.htm
            VECTOR: pCYPAC2
            IMPORTANT: This sequence is not the entire insert of clone
            RP4-809E13 It may be shorter because we sequence overlapping
            sections only once, except for a short overlap.
            The true right end of clone RP4-809E13 is at 67201 in this
            sequence. The true right end of clone RP3-473B4 is at 100 in this
            sequence.
FEATURES             Location/Qualifiers
     source          1. .67201
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="X"
                     /map="q26.1-27.1"
                     /clone="RP4-809E13"
                     /clone_lib="RPCI-4"
     misc_feature    1. .159
                     /note="match: GSS: Em:B42366"
     misc_feature    complement(1. .74)
                     /note="match: STS: Em:AL008947 Em:HS473B4T"
     repeat_region   550. .597
                     /note="2 copies 24 mer 95% conserved"
     repeat_region   758. .922
                     /note="L1ME repeat: matches 5531. .5695 of consensus"
     repeat_region   939. .1213
                     /note="L2 repeat: matches 1806. .2096 of consensus"
     repeat_region   1271. .1565
                     /note="AluSx repeat: matches 1. .295 of consensus"
     repeat_region   2011. .2163
                     /note="L1M4 repeat: matches 3743. .3917 of consensus"
     repeat_region   2642. .2689
                     /note="2 copies 24 mer 97% conserved"
     repeat_region   3206. .3274
                     /note="3 copies 23 mer 81% conserved"
     repeat_region   3243. .3418
                     /note="88 copies 2 mer aa 56% conserved"
     repeat_region   3534. .3636
                     /note="MIR repeat: matches 16. .115 of consensus"
     repeat_region   4253. .4518
                     /note="AluJo repeat: matches 9. .273 of consensus"
     repeat_region   5225. .5524
                     /note="AluSx repeat: matches 1. .299 of consensus"
     repeat_region   7872. .8066
                     /note="MER67C repeat: matches 1. .200 of consensus"
     misc_feature    complement(11087. .11278)
                     /note="match: GSS: Em:AQ283138"
     repeat_region   14329. .14451
                     /note="FLAM_C repeat: matches 1. .126 of consensus"
     repeat_region   15570. .15671
                     /note="MIR repeat: matches 110. .218 of consensus"
     misc_feature    17271. .17692
                     /note="match: GSS: Em:AQ182699"
     misc_feature    17275. .17984
                     /note="match: GSS: Em:AQ553839"
     repeat_region   18475. .18708
                     /note="L1PA5 repeat: matches 5912. .6141 of consensus"
     repeat_region   18711. .18752
                     /note="21 copies 2 mer aa 78% conserved"
     misc_feature    18865. .19171
                     /note="match: STS: Em:AA894450"
     repeat_region   19099. .19303
                     /note="LTR37B repeat: matches 256. .464 of consensus"
     gene            complement(19859. .20487)
                     /gene="dJ809E13.1"
                     /pseudo
     CDS             complement(19859. .20487)
                     /gene="dJ809E13.1"
                     /note="dJ809E13.1 (putative prenylated protein pseudogene)
                     match: cDNAs: Em:AB019527 Em:AF086537 Em:L76224 Em:U77782
                     match: ESTs: Em:AI815482 Em:AI573275 Em:AA659153 Em:N72726
                     Em:AI183679 Em:AW370707 Em:AW370768
                     match: proteins: Sw:P20824 Tr:O88409 Sw:P11675 Tr:O95751
                     Tr:Q15779 Tr:Q9X8Z2 Tr:O70555 Tr:Q61824 Tr:Q63729
                     Tr:Q9Y172 Tr:O68872 Tr:O15255 Tr:Q14468 Tr:Q9YDA0
```

```
                Sw:P33479 Sw:Q03391"
                /pseudo
                /codon_start=1
                /evidence=not_experimental
misc_feature    20312. .20917
                /note="CpG island"
                /evidence=not_experimental
repeat_region   21213. .21450
                /note="L1MC4 repeat: matches 7288. .7525 of consensus"
repeat_region   21645. .21701
                /note="L1 repeat: matches 3880. .3935 of consensus"
repeat_region   21704. .21981
                /note="MLT1D repeat: matches 260. .568 of consensus"
repeat_region   21963. .22196
                /note="MLT1D repeat: matches 73. .323 of consensus"
repeat_region   22197. .23067
                /note="MER11D repeat: matches 1. .897 of consensus"
repeat_region   23068. .23143
                /note="MLT1D repeat: matches 1. .73 of consensus"
repeat_region   23154. .23260
                /note="L1 repeat: matches 3759. .3870 of consensus"
repeat_region   23261. .23532
                /note="L1ME2 repeat: matches 5757. .6048 of consensus"
repeat_region   23510. .23715
                /note="L1MC/D repeat: matches 5411. .5596 of consensus"
repeat_region   23816. .24409
                /note="L1MC/D repeat: matches 4796. .5397 of consensus"
repeat_region   24410. .24486
                /note="MER2 repeat: matches 268. .345 of consensus"
repeat_region   24496. .24630
                /note="L1 repeat: matches 4189. .4323 of consensus"
repeat_region   24643. .24784
                /note="L1P repeat: matches 3712. .3855 of consensus"
repeat_region   24774. .25830
                /note="L1PB2 repeat: matches 4333. .5439 of consensus"
repeat_region   25828. .25903
                /note="L1P1 repeat: matches 5642. .5717 of consensus"
repeat_region   25904. .26345
                /note="L1HS repeat: matches 5705. .6146 of consensus"
repeat_region   26366. .27087
                /note="L1PB2 repeat: matches 5423. .6155 of consensus"
repeat_region   27113. .27322
                /note="MER2 repeat: matches 1. .212 of consensus"
repeat_region   27323. .28033
                /note="L1M4 repeat: matches 4063. .4797 of consensus"
repeat_region   28034. .28237
                /note="L1PA12 repeat: matches 5944. .6158 of consensus"
repeat_region   28238. .28361
                /note="L1M4 repeat: matches 3943. .4063 of consensus"
repeat_region   28361. .28536
                /note="L1MC/D repeat: matches 5613. .5792 of consensus"
repeat_region   28537. .28689
                /note="MLT1E repeat: matches 1. .174 of consensus"
repeat_region   28690. .29650
                /note="LTR5 repeat: matches 1. .969 of consensus"
repeat_region   29651. .29765
                /note="MLT1E repeat: matches 174. .282 of consensus"
repeat_region   29848. .30076
                /note="L1MA2 repeat: matches 6073. .6226 of consensus"
repeat_region   30077. .30137
                /note="L1MA1 repeat: matches 6169. .6226 of consensus"
misc_feature    30383. .31300
                /note="CpG island"
                /evidence=not_experimental
gene            30658. .31891
                /gene="dJ809E13.2"
mRNA            30658. .31891
                /gene="dJ809E13.2"
                /product="dJ809E13.2 (putative prenylated protein)"
                /note="match: cDNAs: Em:AF052096 Em:AF038168 Em:Y13374
                match: ESTs: Em:AI742011 Em:F21403 Em:W76590 Em:AA307887
                Em:F21483 Em:N64112 Em:AA421975 Em:AA031541 Em:AA122144
                Em:AW410876 Em:AW370386 Em:F37666 Em:AI027000 Em:AA091701
                Em:AA302562 Em:AA070647 Em:AA733171 Em:AA076516 Em:C04431
                Em:AA864654 Em:AI339190 Em:AI086752 Em:AI346926

  Query Match             84.8%;  Score 17.8;  DB 9;  Length 67201;
  Best Local Similarity   90.5%;  Pred. No. 4.8e+02;
  Matches   19;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy          1 AAGCCTGATTGATGACCAGGG 21
              || |||| ||||||||||||||
Db      19111 AAACCTGGTTGATGACCAGGG 19091


RESULT 8
AC102710
LOCUS       AC102710               163746 bp    DNA     linear   HTG 17-FEB-2003
DEFINITION  Mus musculus clone RP24-160E14, WORKING DRAFT SEQUENCE, 9 unordered
            pieces.
ACCESSION   AC102710
VERSION     AC102710.3  GI:28395973
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 163746)
  AUTHORS   Birren,B., Nusbaum,C. and Lander,E.
  TITLE     Mus musculus, clone RP24-160E14
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 163746)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Ali,A., Allen,N.,
            Anderson,S., Barna,N., Bastien,V., Boguslavkiy,L., Boukhgalter,B.,
            Brown,A., Camarata,J., Campopiano,A., Chang,J., Chazaro,B.,
            Choepel,Y., Colangelo,M., Collins,S., Collymore,A., Cook,A.,
            Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S., Faro,S.,
            Ferreira,P., FitzHugh,W., Gage,D., Galagan,J., Gardyna,S.,
            Ginde,S., Gord,S., Goyette,M., Graham,L., Grand-Pierre,N.,
            Hagos,B., Heaford,A., Horton,L., Hulme,W., Iliev,I., Johnson,R.,
            Jones,C., Kamat,A., Karatas,A., Kells,C., LaRocque,K.,
            Lamazares,R., Landers,T., Lehoczky,J., Levine,R., Liu,G.,
            MacLean,C., Macdonald,P., Major,J., Marquis,N., Matthews,C.,
            McCarthy,M., McEwan,P., McKernan,K., McPheeters,R., Meldrim,J.,
            Meneus,L., Mihova,T., Mlenga,V., Murphy,T., Naylor,J., Nguyen,C.,
            Norbu,C., Norman,C.H., O'Connor,T., O'Donnell,P., O'Neil,D.,
            Oliver,J., Peterson,K., Phunkhang,P., Pierre,N., Pollara,V.,
            Raymond,C., Retta,R., Rieback,M., Riley,R., Rise,C., Rogov,P.,
            Roman,J., Rosetti,M., Roy,A., Santos,R., Schauer,S., Schupback,R.,
            Seaman,S., Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N.,
            Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,
            Topham,K., Travers,M., Travis,N., Trigilio,J., Vassiliev,H.,
            Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G.,
            Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (23-NOV-2001) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
REFERENCE   3  (bases 1 to 163746)
  AUTHORS   Birren,B., Nusbaum,C., Lander,E., Abouelleil,A., Allen,N.,
            Anderson,S., Arachchi,H.M., Barna,N., Bastien,V., Bloom,T.,
            Boguslavkiy,L., Boukhgalter,B., Camarata,J., Chang,J., Choepel,Y.,
            Collymore,A., Cook,A., Cooke,P., Corum,B., DeArellano,K.,
            Diaz,J.S., Dodge,S., Dooley,K., Dorris,L., Erickson,J., Faro,S.,
            Ferreira,P., FitzGerald,M., Gage,D., Galagan,J., Gardyna,S.,
            Graham,L., Grand-Pierre,N., Hafez,N., Hagopian,D., Hagos,B.,
            Hall,J., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C.,
            Kamat,A., Karatas,A., Kells,C., Landers,T., Levine,R.,
            Lindblad-Toh,K., Liu,G., Lui,A., Mabbitt,R., MacLean,C.,
            Macdonald,P., Major,J., Manning,J., Matthews,C., McCarthy,M.,
            Meldrim,J., Meneus,L., Mihova,T., Mlenga,V., Murphy,T., Naylor,J.,
            Nguyen,C., Nicol,R., Norbu,C., O'Connor,T., O'Donnell,P.,
            O'Neil,D., Oliver,J., Peterson,K., Phunkhang,P., Pierre,N.,
            Rachupka,A., Ramasamy,U., Raymond,C., Retta,R., Rise,C., Rogov,P.,
            Roman,J., Schauer,S., Schupback,R., Seaman,S., Severy,P., Smith,C.,
            Spencer,B., Stange-Thomann,N., Stojanovic,N., Stubbs,M.,
            Talamas,J., Tesfaye,S., Theodore,J., Topham,K., Travers,M.,
```

```
            Vassiliev,H., Venkataraman,V.S., Viel,R., Vo,A., Wilson,B., Wu,X.,
            Wyman,D., Young,G., Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (17-FEB-2003) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT     On Feb 17, 2003 this sequence version replaced gi:22381702.
            All repeats were identified using RepeatMasker:
            Smit, A.F.A. & Green, P. (1996-1997)
            http://ftp.genome.washington.edu/RM/RepeatMasker.html
            -------------- Genome Center
                Center: Whitehead Institute/ MIT Center for Genome Research
                Center code: WIBR
                Web site: http://www-seq.wi.mit.edu
                Contact: sequence_submissions@genome.wi.mit.edu
            -------------- Project Information
                Center project name: L19331
                Center clone name: 160_E_14
            -------------- Summary Statistics
                Sequencing vector: Plasmid; n/a; 100% of reads
                Chemistry: Dye-terminator Big Dye; 100% of reads
                Assembly program: Phrap; version 0.960731
                Consensus quality: 162090 bases at least Q40
                Consensus quality: 162587 bases at least Q30
                Consensus quality: 162779 bases at least Q20
                Insert size: 157000; agarose-fp
                Insert size: 162946; sum-of-contigs
                Quality coverage: 9.4 in Q20 bases; agarose-fp
                Quality coverage: 9.0 in Q20 bases; sum-of-contigs
            --------------.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 9 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1    31737: contig of 31737 bp in length
            *    31738    31837: gap of 100 bp
            *    31838    36585: contig of 4748 bp in length
            *    36586    36685: gap of 100 bp
            *    36686    40932: contig of 4247 bp in length
            *    40933    41032: gap of 100 bp
            *    41033    44827: contig of 3795 bp in length
            *    44828    44927: gap of 100 bp
            *    44928    60759: contig of 15832 bp in length
            *    60760    60859: gap of 100 bp
            *    60860    73424: contig of 12565 bp in length
            *    73425    73524: gap of 100 bp
            *    73525   116324: contig of 42800 bp in length
            *   116325   116424: gap of 100 bp
            *   116425   142656: contig of 26232 bp in length
            *   142657   142756: gap of 100 bp
            *   142757   163746: contig of 20990 bp in length.
FEATURES             Location/Qualifiers
     source          1. .163746
                     /organism="Mus musculus"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:10090"
                     /clone="RP24-160E14"
                     /clone_lib="RPCI-24 Male Mouse BAC"
     misc_feature    1. .31737
                     /note="assembly_fragment
                     clone_end:SP6
                     vector_side:left"
     misc_feature    31838. .36585
                     /note="assembly_fragment"
     misc_feature    36686. .40932
                     /note="assembly_fragment"
     misc_feature    41033. .44827
                     /note="assembly_fragment"
     misc_feature    44928. .60759
                     /note="assembly_fragment"
     misc_feature    60860. .73424
                     /note="assembly_fragment"
     misc_feature    73525. .116324
                     /note="assembly_fragment"
     misc_feature    116425. .142656
                     /note="assembly_fragment"
     misc_feature    142757. .163746
                     /note="assembly_fragment
                     clone_end:T7
                     vector_side:right"
BASE COUNT    44170 a  38132 c  38033 g  42611 t    800 others
ORIGIN

  Query Match             84.8%;  Score 17.8;  DB 2;  Length 163746;
  Best Local Similarity   90.5%;  Pred. No. 4.7e+02;
  Matches   19;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy          1 AAGCCTGATTGATGACCAGGG 21
              |||| ||| ||||||||||||
Db     159477 AAGCTTGAATGATGACCAGGG 159497


RESULT 9
AC109462/c
LOCUS       AC109462               174191 bp    DNA     linear   PRI 06-MAR-2002
DEFINITION  Homo sapiens chromosome 16 clone RP11-26L20, complete sequence.
ACCESSION   AC109462 AC016225
VERSION     AC109462.2  GI:19172840
KEYWORDS    HTG.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 174191)
  AUTHORS   DOE Joint Genome Institute.
  TITLE     Sequencing of Human Chromosome 16
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 174191)
  AUTHORS   DOE Joint Genome Institute.
  TITLE     Direct Submission
  JOURNAL   Submitted (04-FEB-2002) Production Sequencing Facility, DOE Joint
            Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
REFERENCE   3  (bases 1 to 174191)
  AUTHORS   DOE Joint Genome Institute.
  TITLE     Direct Submission
  JOURNAL   Submitted (06-MAR-2002) Production Sequencing Facility, DOE Joint
            Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
COMMENT     On or before Mar 6, 2002 this sequence version replaced
            gi:10198460, gi:18483458.
            Sequence Quality Assessment:
             This entry has been annotated with sequence quality
             estimates computed by the Phrap assembly program.
             All manually edited bases have been reduced to quality zero.
             Quality levels above 40 are expected to have less than
             1 error in 10,000 bp.
             Base-by-base quality values are not generally visible from the
             GenBank flat file format but are available as part
             of this entry's ASN.1 file.
            ----------------------.
            Sequence Quality Assessment:
             This entry has been annotated with sequence quality
             estimates computed by the Phrap assembly program.
             All manually edited bases have been reduced to quality zero.
             Quality levels above 40 are expected to have less than
             1 error in 10,000 bp.
             Base-by-base quality values are not generally visible from the
             GenBank flat file format but are available as part
             of this entry's ASN.1 file.
            ----------------------.
FEATURES             Location/Qualifiers
     source          1. .174191
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
```

```
                     /db_xref="taxon:9606"
                     /chromosome="16"
                     /clone="RP11-26L20"
BASE COUNT    51311 a  36924 c  37255 g  48701 t
ORIGIN

  Query Match             84.8%;  Score 17.8;  DB 9;  Length 174191;
  Best Local Similarity   90.5%;  Pred. No. 4.7e+02;
  Matches   19;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy          1 AAGCCTGATTGATGACCAGGG 21
              ||||||||||  ||  |||||||
Db      39604 AAGCCTGATTTATCACCAGGG 39584


RESULT 10
AC113908/c
LOCUS       AC113908              177733 bp    DNA     linear   HTG 15-NOV-2002
DEFINITION  Rattus norvegicus clone CH230-396P4, WORKING DRAFT SEQUENCE.
ACCESSION   AC113908
VERSION     AC113908.9  GI:25008249
KEYWORDS    HTG; HTGS_PHASE2; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE      Rattus norvegicus (Norway rat)
  ORGANISM  Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Rattus.
REFERENCE   1  (bases 1 to 177733)
  AUTHORS   Muzny,D.Marie., Metzker,M.Lee., Abramzon,S., Adams,C., Alder,J.,
            Allen,C., Allen,H., Alsbrooks,S., Amin,A., Anguiano,D.,
            Anyalebechi,V., Aoyagi,A., Ayodeji,M., Baca,E., Baden,H.,
            Baldwin,D., Bandaranaike,D., Barber,M., Barnstead,M., Benahmed,F.,
            Biswalo,K., Blair,J., Blankenburg,K., Blyth,P., Brown,M.,
            Bryant,N., Buhay,C., Burch,P., Burrell,K., Calderon,E.,
            Cardenas,V., Carter,K., Cavazos,I., Ceasar,H., Center,A.,
            Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Y., Chen,Z., Chu,J.,
            Cleveland,C., Cockrell,R., Cox,C., Coyle,M., Cree,A., D'Souza,L.,
            Davila,M.L., Davis,C., Davy-Carroll,L., De Anda,C., Dederich,D.,
            Delgado,O., Denson,S., Deramo,C., Ding,Y., Dinh,H., Divya,K.,
            Draper,H., Dugan-Rocha,S., Dunn,A., Durbin,K., Duval,B., Eaves,K.,
            Egan,A., Escotto,M., Eugene,C., Evans,C.A., Falls,T., Fan,G.,
            Fernandez,S., Finley,M., Flagg,N., Forbes,L., Foster,M., Foster,P.,
            Fraser,C.M., Gabisi,A., Ganta,R., Garcia,A., Garner,T., Garza,M.,
            Gebregeorgis,E., Geer,K., Gill,R., Grady,M., Guerra,W., Guevara,W.,
            Gunaratne,P., Haaland,W., Hamil,C., Hamilton,C., Hamilton,K.,
            Harvey,Y., Havlak,P., Hawes,A., Henderson,N., Hernandez,J.,
            Hernandez,R., Hines,S., Hladun,S.L., Hodgson,A., Hogues,M.,
            Hollins,B., Howells,S., Hulyk,S., Hume,J., Idlebird,D., Jackson,A.,
            Jackson,L., Jacob,L., Jiang,H., Johnson,B., Johnson,R., Jolivet,A.,
            Karpathy,S., Kelly,S., Kelly,S., Khan,Z., King,L., Kovar,C.,
            Kowis,C., Kraft,C.L., Lebow,H., Levan,J., Lewis,L., Li,Z., Liu,J.,
            Liu,J., Liu,W., Liu,Y., London,P., Longacre,S., Lopez,J.,
            Lorensuhewa,L., Loulseged,H., Lozado,R.J., Lu,X., Ma,J.,
            Maheshwari,M., Mahindartne,M., Mahmoud,M., Malloy,K., Mangum,A.,
            Mangum,B., Mapua,P., Martin,K., Martin,R., Martinez,E.,
            Mawhiney,S., McLeod,M.P., McNeill,T.Z., Meenen,E.,
            Milosavljevic,A., Miner,G., Minja,E., Montemayor,J., Moore,S.,
            Morgan,M., Morris,K., Morris,S., Munidasa,M., Murphy,M., Nair,L.,
            Nankervis,C., Neal,D., Newton,N., Nguyen,N., Norris,S.,
            Nwaokelemeh,O., Okwuonu,G., Olarnpunsagoon,A., Pal,S., Parks,K.,
            Pasternak,S., Paul,H., Perez,A., Perez,L., Pfannkoch,C.,
            Plopper,F., Poindexter,A., Popovic,D., Primus,E., Pu,L.-L.,
            Puazo,M., Quiroz,J., Rachlin,E., Reeves,K., Regier,M.A., Reigh,R.,
            Reilly,B., Reilly,M., Ren,Y., Reuter,M., Richards,S., Riggs,F.,
            Rives,C., Rodkey,T., Rojas,A., Rose,M., Rose,R., Ruiz,S.J.,
            Sanders,W., Savery,G., Scherer,S., Scott,G., Shatsman,S., Shen,H.,
            Shetty,J., Shvartsbeyn,A., Sisson,I., Sitter,C.D., Smajs,D.,
            Sneed,A., Sodergren,E., Song,X.-Z., Sorelle,R., Sosa,J.,
            Steimle,M., Strong,R., Sutton,A., Svatek,A., Tabor,P., Taylor,C.,
            Taylor,T., Thomas,N., Thomas,S., Tingey,A., Trejos,Z., Usmani,K.,
            Valas,R., Vera,V., Villasana,D., Waldron,L., Walker,B., Wang,J.,
            Wang,Q., Wang,S., Warren,J., Warren,R., Wei,X., White,F.,
            Williams,G., Willson,R., Wleczyk,R., Wooden,H., Worley,K.,
            Wright,D., Wright,R., Wu,J., Yakub,S., Yen,J., Yoon,L., Yoon,V.,
            Yu,F., Zhang,J., Zhou,J., Zhou,X., Zhao,S., Dunn,D., von
            Niederhausern,A., Weiss,R., Smith,D.R., Holt,R.A., Smith,H.O.,
            Weinstock,G. and Gibbs,R.A.
  TITLE     Direct Submission
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 177733)
  AUTHORS   Worley,K.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (05-MAR-2002) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
REFERENCE   3  (bases 1 to 177733)
  AUTHORS   Rat Genome Sequencing Consortium.
  TITLE     Direct Submission
  JOURNAL   Submitted (15-NOV-2002) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
COMMENT     On Nov 15, 2002 this sequence version replaced gi:23815563.
            The sequence in this assembly is a combination of BAC based reads
            and whole genome shotgun sequencing reads assembled using Atlas
            (http://www.hgsc.bcm.tmc.edu/projects/rat/). Each contig described
            in the feature table below represents a scaffold in the Atlas
            assembly (a 'contig-scaffold'). Within each contig-scaffold,
            individual sequence contigs are ordered and oriented, and separated
            by sized gaps filled with Ns to the estimated size. The sequence
            may extend beyond the ends of the clone and there may be sequence
            contigs within a contig-scaffold that consist entirely of whole
            genome shotgun sequence reads. Both end sequences and whole genome
            shotgun sequence only contigs will be indicated in the feature
            table.
            -------------- Genome Center
                Center: Baylor College of Medicine
                Center code: BCM
                Web site: http://www.hgsc.bcm.tmc.edu/
                Contact: hgsc-help@bcm.tmc.edu
            -------------- Project Information
                Center project name: GTDH
                Center clone name: CH230-396P4
            -------------- Summary Statistics
                Assembly program: Phrap; version 0.990329
                Consensus quality: 165823 bases at least Q40
                Consensus quality: 167723 bases at least Q30
                Consensus quality: 168770 bases at least Q20
                Estimated insert size: 172340; sum-of-contigs estimation
                Quality coverage: 7x in Q20 bases; sum-of-contigs estimation
            --------------
            * NOTE: Estimated insert size may differ from sequence length
            *   (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 1 contigs. Gaps between the contigs
            * are represented as runs of N. The order of the pieces
            * is believed to be correct as given, however the sizes
            * of the gaps between them are based on estimates that have
            * provided by the submittor.
            * This sequence will be replaced
            * by the finished sequence as soon as it is available and
            * the accession number will be preserved.
            *       1   177733: contig of 177733 bp in length.
FEATURES             Location/Qualifiers
     source          1. .177733
                     /organism="Rattus norvegicus"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:10116"
                     /clone="CH230-396P4"
     misc_feature    342. .1213
                     /note="clone_boundary
                     clone_end:T7
                     site:
                     end_sequence:BZ175527"
     misc_feature    complement(176718. .177330)
                     /note="clone_boundary
                     clone_end:Sp6
```

```
                    site:
                    end_sequence:BZ175528"
BASE COUNT    48653 a  37788 c  36881 g  46258 t   8153 others
ORIGIN

  Query Match             84.8%;  Score 17.8;  DB 2;  Length 177733;
  Best Local Similarity   90.5%;  Pred. No. 4.7e+02;
  Matches   19;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy          1 AAGCCTGATTGATGACCAGGG 21
              ||||||||||||||| ||| |
Db      72763 AAGCCTGATTGATGATCAGAG 72743


RESULT 11
AC128288/c
LOCUS       AC128288              180353 bp    DNA     linear   HTG 20-NOV-2002
DEFINITION  Rattus norvegicus clone CH230-390F9, WORKING DRAFT SEQUENCE.
ACCESSION   AC128288
VERSION     AC128288.3  GI:25137732
KEYWORDS    HTG; HTGS_PHASE2; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE      Rattus norvegicus (Norway rat)
  ORGANISM  Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Rattus.
REFERENCE   1  (bases 1 to 180353)
  AUTHORS   Muzny,D.Marie., Metzker,M.Lee., Abramzon,S., Adams,C., Alder,J.,
            Allen,C., Allen,H., Alsbrooks,S., Amin,A., Anguiano,D.,
            Anyalebechi,V., Aoyagi,A., Ayodeji,M., Baca,E., Baden,H.,
            Baldwin,D., Bandaranaike,D., Barber,M., Barnstead,M., Benahmed,F.,
            Biswalo,K., Blair,J., Blankenburg,K., Blyth,P., Brown,M.,
            Bryant,N., Buhay,C., Burch,P., Burrell,K., Calderon,E.,
            Cardenas,V., Carter,K., Cavazos,I., Ceasar,H., Center,A.,
            Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Y., Chen,Z., Chu,J.,
            Cleveland,C., Cockrell,R., Cox,C., Coyle,M., Cree,A., D'Souza,L.,
            Davila,M.L., Davis,C., Davy-Carroll,L., De Anda,C., Dederich,D.,
            Delgado,O., Denson,S., Deramo,C., Ding,Y., Dinh,H., Divya,K.,
            Draper,H., Dugan-Rocha,S., Dunn,A., Durbin,K., Duval,B., Eaves,K.,
            Egan,A., Escotto,M., Eugene,C., Evans,C.A., Falls,T., Fan,G.,
            Fernandez,S., Finley,M., Flagg,N., Forbes,L., Foster,M., Foster,P.,
            Fraser,C.M., Gabisi,A., Ganta,R., Garcia,A., Garner,T., Garza,M.,
            Gebregeorgis,E., Geer,K., Gill,R., Grady,M., Guerra,W., Guevara,W.,
            Gunaratne,P., Haaland,W., Hamil,C., Hamilton,C., Hamilton,K.,
            Harvey,Y., Havlak,P., Hawes,A., Henderson,N., Hernandez,J.,
            Hernandez,R., Hines,S., Hladun,S.L., Hodgson,A., Hogues,M.,
            Hollins,B., Howells,S., Hulyk,S., Hume,J., Idlebird,D., Jackson,A.,
            Jackson,L., Jacob,L., Jiang,H., Johnson,B., Johnson,R., Jolivet,A.,
            Karpathy,S., Kelly,S., Kelly,S., Khan,Z., King,L., Kovar,C.,
            Kowis,C., Kraft,C.L., Lebow,H., Levan,J., Lewis,L., Li,Z., Liu,J.,
            Liu,J., Liu,W., Liu,Y., London,P., Longacre,S., Lopez,J.,
            Lorensuhewa,L., Loulseged,H., Lozado,R.J., Lu,X., Ma,J.,
            Maheshwari,M., Mahindartne,M., Mahmoud,M., Malloy,K., Mangum,A.,
            Mangum,B., Mapua,P., Martin,K., Martin,R., Martinez,E.,
            Mawhiney,S., McLeod,M.P., McNeill,T.Z., Meenen,E.,
            Milosavljevic,A., Miner,G., Minja,E., Montemayor,J., Moore,S.,
            Morgan,M., Morris,K., Morris,S., Munidasa,M., Murphy,M., Nair,L.,
            Nankervis,C., Neal,D., Newton,N., Nguyen,N., Norris,S.,
            Nwaokelemeh,O., Okwuonu,G., Olarnpunsagoon,A., Pal,S., Parks,K.,
            Pasternak,S., Paul,H., Perez,A., Perez,L., Pfannkoch,C.,
            Plopper,F., Poindexter,A., Popovic,D., Primus,E., Pu,L.-L.,
            Puazo,M., Quiroz,J., Rachlin,E., Reeves,K., Regier,M.A., Reigh,R.,
            Reilly,B., Reilly,M., Ren,Y., Reuter,M., Richards,S., Riggs,F.,
            Rives,C., Rodkey,T., Rojas,A., Rose,M., Rose,R., Ruiz,S.J.,
            Sanders,W., Savery,G., Scherer,S., Scott,G., Shatsman,S., Shen,H.,
            Shetty,J., Shvartsbeyn,A., Sisson,I., Sitter,C.D., Smajs,D.,
            Sneed,A., Sodergren,E., Song,X.-Z., Sorelle,R., Sosa,J.,
            Steimle,M., Strong,R., Sutton,A., Svatek,A., Tabor,P., Taylor,C.,
            Taylor,T., Thomas,N., Thomas,S., Tingey,A., Trejos,Z., Usmani,K.,
            Valas,R., Vera,V., Villasana,D., Waldron,L., Walker,B., Wang,J.,
            Wang,Q., Wang,S., Warren,J., Warren,R., Wei,X., White,F.,
            Williams,G., Willson,R., Wleczyk,R., Wooden,H., Worley,K.,
            Wright,D., Wright,R., Wu,J., Yakub,S., Yen,J., Yoon,L., Yoon,V.,
            Yu,F., Zhang,J., Zhou,J., Zhou,X., Zhao,S., Dunn,D., von
            Niederhausern,A., Weiss,R., Smith,D.R., Holt,R.A., Smith,H.O.,
            Weinstock,G. and Gibbs,R.A.
  TITLE     Direct Submission
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 180353)
  AUTHORS   Worley,K.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (19-JUL-2002) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
REFERENCE   3  (bases 1 to 180353)
  AUTHORS   Rat Genome Sequencing Consortium.
  TITLE     Direct Submission
  JOURNAL   Submitted (20-NOV-2002) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
COMMENT     On Nov 20, 2002 this sequence version replaced gi:22856104.
            The sequence in this assembly is a combination of BAC based reads
            and whole genome shotgun sequencing reads assembled using Atlas
            (http://www.hgsc.bcm.tmc.edu/projects/rat/). Each contig described
            in the feature table below represents a scaffold in the Atlas
            assembly (a 'contig-scaffold'). Within each contig-scaffold,
            individual sequence contigs are ordered and oriented, and separated
            by sized gaps filled with Ns to the estimated size. The sequence
            may extend beyond the ends of the clone and there may be sequence
            contigs within a contig-scaffold that consist entirely of whole
            genome shotgun sequence reads. Both end sequences and whole genome
            shotgun sequence only contigs will be indicated in the feature
            table.
            -------------- Genome Center
                Center: Baylor College of Medicine
                Center code: BCM
                Web site: http://www.hgsc.bcm.tmc.edu/
                Contact: hgsc-help@bcm.tmc.edu
            -------------- Project Information
                Center project name: GYDD
                Center clone name: CH230-390F9
            -------------- Summary Statistics
                Assembly program: Phrap; version 0.990329
                Consensus quality: 168606 bases at least Q40
                Consensus quality: 170052 bases at least Q30
                Consensus quality: 171200 bases at least Q20
                Estimated insert size: 171038; sum-of-contigs estimation
                Quality coverage: 8x in Q20 bases; sum-of-contigs estimation
            --------------
            * NOTE: Estimated insert size may differ from sequence length
            *   (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 1 contigs. Gaps between the contigs
            * are represented as runs of N. The order of the pieces
            * is believed to be correct as given, however the sizes
            * of the gaps between them are based on estimates that have
            * provided by the submittor.
            * This sequence will be replaced
            * by the finished sequence as soon as it is available and
            * the accession number will be preserved.
            *       1   180353: contig of 180353 bp in length.
FEATURES             Location/Qualifiers
     source          1. .180353
                     /organism="Rattus norvegicus"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:10116"
                     /clone="CH230-390F9"
     misc_feature    1. .1070
                     /note="wgs_contig"
     misc_feature    2565. .4033
                     /note="wgs_contig"
     misc_feature    4915. .6314
                     /note="wgs_contig"
BASE COUNT    48322 a  37960 c  37295 g  48481 t   8295 others
ORIGIN
```

```
  Query Match             84.8%;  Score 17.8;  DB 2;  Length 180353;
  Best Local Similarity   90.5%;  Pred. No. 4.7e+02;
  Matches   19;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy          1 AAGCCTGATTGATGACCAGGG 21
              ||||||||||||||| ||| |
Db      30172 AAGCCTGATTGATGATCAGAG 30152


RESULT 12
AC117856/c
LOCUS       AC117856               186118 bp    DNA     linear   HTG 11-OCT-2002
DEFINITION  Rattus norvegicus clone CH230-327H18, WORKING DRAFT SEQUENCE, 3
            unordered pieces.
ACCESSION   AC117856
VERSION     AC117856.4  GI:23814006
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE      Rattus norvegicus (Norway rat)
  ORGANISM  Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Rattus.
REFERENCE   1  (bases 1 to 186118)
  AUTHORS   Muzny,D.Marie., Metzker,M.Lee., Abramzon,S., Adams,C., Alder,J.,
            Allen,C., Allen,H., Alsbrooks,S., Amin,A., Anguiano,D.,
            Anyalebechi,V., Aoyagi,A., Ayodeji,M., Baca,E., Baden,H.,
            Baldwin,D., Bandaranaike,D., Barber,M., Barnstead,M., Benahmed,F.,
            Biswalo,K., Blair,J., Blankenburg,K., Blyth,P., Brown,M.,
            Bryant,N., Buhay,C., Burch,P., Burrell,K., Calderon,E.,
            Cardenas,V., Carter,K., Cavazos,I., Ceasar,H., Center,A.,
            Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Y., Chen,Z., Chu,J.,
            Cleveland,C., Cockrell,R., Cox,C., Coyle,M., Cree,A., D'Souza,L.,
            Davila,M.L., Davis,C., Davy-Carroll,L., De Anda,C., Dederich,D.,
            Delgado,O., Denson,S., Deramo,C., Ding,Y., Dinh,H., Divya,K.,
            Draper,H., Dugan-Rocha,S., Dunn,A., Durbin,K., Duval,B., Eaves,K.,
            Egan,A., Escotto,M., Eugene,C., Evans,C.A., Falls,T., Fan,G.,
            Fernandez,S., Finley,M., Flagg,N., Forbes,L., Foster,M., Foster,P.,
            Fraser,C.M., Gabisi,A., Ganta,R., Garcia,A., Garner,T., Garza,M.,
            Gebregeorgis,E., Geer,K., Gill,R., Grady,M., Guerra,W., Guevara,W.,
            Gunaratne,P., Haaland,W., Hamil,C., Hamilton,C., Hamilton,K.,
            Harvey,Y., Havlak,P., Hawes,A., Henderson,N., Hernandez,J.,
            Hernandez,R., Hines,S., Hladun,S.L., Hodgson,A., Hogues,M.,
            Hollins,B., Howells,S., Hulyk,S., Hume,J., Idlebird,D., Jackson,A.,
            Jackson,L., Jacob,L., Jiang,H., Johnson,B., Johnson,R., Jolivet,A.,
            Karpathy,S., Kelly,S., Kelly,S., Khan,Z., King,L., Kovar,C.,
            Kowis,C., Kraft,C.L., Lebow,H., Levan,J., Lewis,L., Li,Z., Liu,J.,
            Liu,J., Liu,W., Liu,Y., London,P., Longacre,S., Lopez,J.,
            Lorensuhewa,L., Loulseged,H., Lozado,R.J., Lu,X., Ma,J.,
            Maheshwari,M., Mahindartne,M., Mahmoud,M., Malloy,K., Mangum,A.,
            Mangum,B., Mapua,P., Martin,K., Martin,R., Martinez,E.,
            Mawhiney,S., McLeod,M.P., McNeill,T.Z., Meenen,E.,
            Milosavljevic,A., Miner,G., Minja,E., Montemayor,J., Moore,S.,
            Morgan,M., Morris,K., Morris,S., Munidasa,M., Murphy,M., Nair,L.,
            Nankervis,C., Neal,D., Newton,N., Nguyen,N., Norris,S.,
            Nwaokelemeh,O., Okwuonu,G., Olarnpunsagoon,A., Pal,S., Parks,K.,
            Pasternak,S., Paul,H., Perez,A., Perez,L., Pfannkoch,C.,
            Plopper,F., Poindexter,A., Popovic,D., Primus,E., Pu,L.-L.,
            Puazo,M., Quiroz,J., Rachlin,E., Reeves,K., Regier,M.A., Reigh,R.,
            Reilly,B., Reilly,M., Ren,Y., Reuter,M., Richards,S., Riggs,F.,
            Rives,C., Rodkey,T., Rojas,A., Rose,M., Rose,R., Ruiz,S.J.,
            Sanders,W., Savery,G., Scherer,S., Scott,G., Shatsman,S., Shen,H.,
            Shetty,J., Shvartsbeyn,A., Sisson,I., Sitter,C.D., Smajs,D.,
            Sneed,A., Sodergren,E., Song,X.-Z., Sorelle,R., Sosa,J.,
            Steimle,M., Strong,R., Sutton,A., Svatek,A., Tabor,P., Taylor,C.,
            Taylor,T., Thomas,N., Thomas,S., Tingey,A., Trejos,Z., Usmani,K.,
            Valas,R., Vera,V., Villasana,D., Waldron,L., Walker,B., Wang,J.,
            Wang,Q., Wang,S., Warren,J., Warren,R., Wei,X., White,F.,
            Williams,G., Willson,R., Wleczyk,R., Wooden,H., Worley,K.,
            Wright,D., Wright,R., Wu,J., Yakub,S., Yen,J., Yoon,L., Yoon,V.,
            Yu,F., Zhang,J., Zhou,J., Zhou,X., Zhao,S., Dunn,D., von
            Niederhausern,A., Weiss,R., Smith,D.R., Holt,R.A., Smith,H.O.,
            Weinstock,G. and Gibbs,R.A.
  TITLE     Direct Submission
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 186118)
  AUTHORS   Worley,K.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (11-APR-2002) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
REFERENCE   3  (bases 1 to 186118)
  AUTHORS   Rat Genome Sequencing Consortium.
  TITLE     Direct Submission
  JOURNAL   Submitted (11-OCT-2002) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
COMMENT     On Oct 11, 2002 this sequence version replaced gi:21745720.
            The sequence in this assembly is a combination of BAC based reads
            and whole genome shotgun sequencing reads assembled using Atlas
            (http://www.hgsc.bcm.tmc.edu/projects/rat/). Each contig described
            in the feature table below represents a scaffold in the Atlas
            assembly (a 'contig-scaffold'). Within each contig-scaffold,
            individual sequence contigs are ordered and oriented, and separated
            by sized gaps filled with Ns to the estimated size. The sequence
            may extend beyond the ends of the clone and there may be sequence
            contigs within a contig-scaffold that consist entirely of whole
            genome shotgun sequence reads. Both end sequences and whole genome
            shotgun sequence only contigs will be indicated in the feature
            table.
            -------------- Genome Center
                Center: Baylor College of Medicine
                Center code: BCM
                Web site: http://www.hgsc.bcm.tmc.edu/
                Contact: hgsc-help@bcm.tmc.edu
            -------------- Project Information
                Center project name: GTTF
                Center clone name: CH230-327H18
            -------------- Summary Statistics
                Assembly program: Phrap; version 0.990329
                Consensus quality: 169552 bases at least Q40
                Consensus quality: 171519 bases at least Q30
                Consensus quality: 172461 bases at least Q20
                Estimated insert size: 173910; sum-of-contigs estimation
                Quality coverage: 7x in Q20 bases; sum-of-contigs estimation
            --------------
            * NOTE: Estimated insert size may differ from sequence length
            *   (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 3 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *       1   182244: contig of 182244 bp in length
            *  182245   182344: gap of unknown length
            *  182345   184242: contig of 1898 bp in length
            *  184243   184342: gap of unknown length
            *  184343   186118: contig of 1776 bp in length.
FEATURES             Location/Qualifiers
     source          1. .186118
                     /organism="Rattus norvegicus"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:10116"
                     /clone="CH230-327H18"
     misc_feature    1. .1031
                     /note="wgs_end_extension
                     clone_end:T7"
     misc_feature    complement(5228. .6123)
                     /note="clone_boundary
                     clone_end:T7
                     site:MboI
                     end_sequence:RXAOL45TJ"
     misc_feature    171082. .171985
                     /note="clone_boundary
```

```
                     clone_end:Sp6
                     site:MboI
                     end_sequence:RXAOL45TV"
     misc_feature    179115. .180519
                     /note="wgs_end_extension
                     clone_end:Sp6"
     misc_feature    181021. .182244
                     /note="wgs_end_extension
                     clone_end:Sp6"
BASE COUNT    49746 a  33747 c  35829 g  53877 t  12919 others
ORIGIN

  Query Match             84.8%;  Score 17.8;  DB 2;  Length 186118;
  Best Local Similarity   90.5%;  Pred. No. 4.7e+02;
  Matches   19;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy          1 AAGCCTGATTGATGACCAGGG 21
              |||||||||| |||||||| |
Db     164995 AAGCCTGATTCATGACCAGTG 164975


RESULT 13
AC111040
LOCUS       AC111040               194454 bp    DNA     linear   HTG 21-AUG-2002
DEFINITION  Mus musculus clone RP24-536C5, WORKING DRAFT SEQUENCE, 11 ordered
            pieces.
ACCESSION   AC111040
VERSION     AC111040.4  GI:22381727
KEYWORDS    HTG; HTGS_PHASE2; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 194454)
  AUTHORS   Birren,B., Nusbaum,C. and Lander,E.
  TITLE     Mus musculus, clone RP24-536C5
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 194454)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Ali,A., Allen,N.,
            Anderson,S., Barna,N., Bastien,V., Boguslavkiy,L., Boukhgalter,B.,
            Brown,A., Camarata,J., Campopiano,A., Chang,J., Chazaro,B.,
            Choepel,Y., Colangelo,M., Collins,S., Collymore,A., Cook,A.,
            Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S., Faro,S.,
            Ferreira,P., FitzHugh,W., Gage,D., Galagan,J., Gardyna,S.,
            Ginde,S., Gord,S., Goyette,M., Graham,L., Grand-Pierre,N.,
            Hagos,B., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C.,
            Kamat,A., Karatas,A., Kells,C., LaRocque,K., Lamazares,R.,
            Landers,T., Lehoczky,J., Levine,R., Liu,G., MacLean,C.,
            Macdonald,P., Major,J., Marquis,N., Matthews,C., McCarthy,M.,
            McEwan,P., McKernan,K., Meldrim,J., Meneus,L., Mihova,T.,
            Mlenga,V., Murphy,T., Naylor,J., Nguyen,C., Nicol,R., Norbu,C.,
            Norman,C.H., O'Connor,T., O'Donnell,P., O'Neil,D., Oliver,J.,
            Peterson,K., Phunkhang,P., Pierre,N., Pollara,V., Raymond,C.,
            Retta,R., Rieback,M., Riley,R., Rise,C., Rogov,P., Roman,J.,
            Rosetti,M., Roy,A., Santos,R., Schauer,S., Schupback,R., Seaman,S.,
            Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N.,
            Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,
            Topham,K., Travers,M., Travis,N., Trigilio,J., Vassiliev,H.,
            Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G.,
            Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (18-FEB-2002) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
REFERENCE   3  (bases 1 to 194454)
  AUTHORS   Birren,B., Nusbaum,C., Lander,E., Ali,A., Allen,N., Anderson,S.,
            Barna,N., Bastien,V., Bloom,T., Boguslavkiy,L., Boukhgalter,B.,
            Camarata,J., Chang,J., Chazaro,B., Choepel,Y., Collymore,A.,
            Cook,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S.,
            Faro,S., Ferreira,P., FitzGerald,M., Gage,D., Galagan,J.,
            Gardyna,S., Gord,S., Graham,L., Grand-Pierre,N., Hagos,B.,
            Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C., Kamat,A.,
            Karatas,A., Kells,C., Landers,T., Levine,R., Lindblad-Toh,K.,
            Liu,G., MacLean,C., Macdonald,P., Major,J., Matthews,C.,
            McCarthy,M., Meldrim,J., Meneus,L., Mihova,T., Mlenga,V.,
            Murphy,T., Naylor,J., Nguyen,C., Nicol,R., Norbu,C., Norman,C.H.,
            O'Connor,T., O'Donnell,P., O'Neil,D., Oliver,J., Peterson,K.,
            Phunkhang,P., Pierre,N., Raymond,C., Retta,R., Rise,C., Rogov,P.,
            Roman,J., Roy,A., Schauer,S., Schupback,R., Seaman,S., Severy,P.,
            Smith,C., Spencer,B., Stange-Thomann,N., Stojanovic,N., Talamas,J.,
            Tesfaye,S., Theodore,J., Topham,K., Travers,M., Vassiliev,H.,
            Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Young,G., Zainoun,J.,
            Zembek,L., Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (21-AUG-2002) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT     On Aug 21, 2002 this sequence version replaced gi:20806337.
            All repeats were identified using RepeatMasker:
            Smit, A.F.A. & Green, P. (1996-1997)
            http://ftp.genome.washington.edu/RM/RepeatMasker.html
            -------------- Genome Center
                Center: Whitehead Institute/ MIT Center for Genome Research
                Center code: WIBR
                Web site: http://www-seq.wi.mit.edu
                Contact: sequence_submissions@genome.wi.mit.edu
            -------------- Project Information
                Center project name: L19423
                Center clone name: 536_C_5
            -------------- Summary Statistics
                Sequencing vector: Plasmid; n/a; 100% of reads
                Chemistry: Dye-terminator Big Dye; 100% of reads
                Assembly program: Phrap; version 0.960731
                Consensus quality: 191374 bases at least Q40
                Consensus quality: 192313 bases at least Q30
                Consensus quality: 192832 bases at least Q20
                Insert size: 182000; agarose-fp
                Insert size: 193454; sum-of-contigs
                Quality coverage: 10.6 in Q20 bases; agarose-fp
                Quality coverage: 10.0 in Q20 bases; sum-of-contigs
            --------------.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 11 contigs. Gaps between the contigs
            * are represented as runs of N. The order of the pieces
            * is believed to be correct as given, however the sizes
            * of the gaps between them are based on estimates that have
            * provided by the submittor.
            * This sequence will be replaced
            * by the finished sequence as soon as it is available and
            * the accession number will be preserved.
            *       1     3686: contig of 3686 bp in length
            *    3687     3786: gap of 100 bp
            *    3787    11203: contig of 7417 bp in length
            *   11204    11303: gap of 100 bp
            *   11304    22194: contig of 10891 bp in length
            *   22195    22294: gap of 100 bp
            *   22295    32225: contig of 9931 bp in length
            *   32226    32325: gap of 100 bp
            *   32326    39995: contig of 7670 bp in length
            *   39996    40095: gap of 100 bp
            *   40096    57751: contig of 17656 bp in length
            *   57752    57851: gap of 100 bp
            *   57852    69354: contig of 11503 bp in length
            *   69355    69454: gap of 100 bp
            *   69455    87475: contig of 18021 bp in length
            *   87476    87575: gap of 100 bp
            *   87576   118208: contig of 30633 bp in length
            *  118209   118308: gap of 100 bp
            *  118309   158787: contig of 40479 bp in length
            *  158788   158887: gap of 100 bp
            *  158888   194454: contig of 35567 bp in length.
FEATURES             Location/Qualifiers
     source          1. .194454
                     /organism="Mus musculus"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:10090"
                     /clone="RP24-536C5"
                     /clone_lib="RPCI-24 Male Mouse BAC"
```

```
     misc_feature    1. .3686
                     /note="assembly_fragment"
     misc_feature    3787. .11203
                     /note="assembly_fragment"
     misc_feature    11304. .22194
                     /note="assembly_fragment"
     misc_feature    22295. .32225
                     /note="assembly_fragment"
     misc_feature    32326. .39995
                     /note="assembly_fragment"
     misc_feature    40096. .57751
                     /note="assembly_fragment"
     misc_feature    57852. .69354
                     /note="assembly_fragment"
     misc_feature    69455. .87475
                     /note="assembly_fragment"
     misc_feature    87576. .118208
                     /note="assembly_fragment"
     misc_feature    118309. .158787
                     /note="assembly_fragment"
     misc_feature    158888. .194454
                     /note="assembly_fragment"
BASE COUNT    53011 a  45075 c  45339 g  50027 t   1002 others
ORIGIN

  Query Match             84.8%;  Score 17.8;  DB 2;  Length 194454;
  Best Local Similarity   90.5%;  Pred. No. 4.7e+02;
  Matches   19;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy          1 AAGCCTGATTGATGACCAGGG 21
              |||| ||| ||||||||||||
Db     134728 AAGCTTGAATGATGACCAGGG 134748


RESULT 14
AL691424/c
LOCUS       AL691424              194520 bp    DNA     linear   ROD 26-JUL-2002
DEFINITION  Mouse DNA sequence from clone RP23-219K12 on chromosome X, complete
            sequence.
ACCESSION   AL691424
VERSION     AL691424.7  GI:22002699
KEYWORDS    HTG.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 194520)
  AUTHORS   Wray,P.
  TITLE     Direct Submission
  JOURNAL   Submitted (26-JUL-2002) Wellcome Trust Sanger Institute, Hinxton,
            Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
            humquery@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
COMMENT     On Jul 29, 2002 this sequence version replaced gi:21912679.
            -------------- Genome Center
            Center: Wellcome Trust Sanger Institute
            Center code: SC
            Web site: http://www.sanger.ac.uk
            Contact: humquery@sanger.ac.uk
            --------------
            During sequence assembly data is compared from overlapping clones.
            Where differences are found these are annotated as variations
            together with a note of the overlapping clone name. Note that the
            variation annotation may not be found in the sequence submission
            corresponding to the overlapping clone, as we submit sequences with
            only a small overlap as described above.
            This sequence was finished as follows unless otherwise noted: all
            regions were either double-stranded or sequenced with an alternate
            chemistry or covered by high quality data (i.e., phred quality >=
            30); an attempt was made to resolve all sequencing problems, such
            as compressions and repeats; all regions were covered by at least
            one plasmid subclone or more than one M13 subclone; and the
            assembly was confirmed by restriction digest. The following
            abbreviations are used to associate primary accession numbers given
            in the feature table with their source databases: Em:, EMBL; Sw:,
            SWISSPROT; Tr:, TREMBL; Wp:, WORMPEP; Information on the WORMPEP
            database can be found at
            http://www.sanger.ac.uk/Projects/C_elegans/wormpep RP23-219K12 is
            from the RPCI-23 Mouse PAC Library
            constructed by the group of Pieter de Jong.
            For further details see http://www.chori.org/bacpac/home.htm
            VECTOR: pBACe3.6.
FEATURES             Location/Qualifiers
     source          1. .194520
                     /organism="Mus musculus"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:10090"
                     /chromosome="X"
                     /clone="RP23-219K12"
                     /clone_lib="RPCI-23"
BASE COUNT    55728 a  38763 c  37184 g  62845 t
ORIGIN

  Query Match             84.8%;  Score 17.8;  DB 10;  Length 194520;
  Best Local Similarity   90.5%;  Pred. No. 4.7e+02;
  Matches   19;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy          1 AAGCCTGATTGATGACCAGGG 21
              | |||||||| ||||||||||
Db      42410 ATGCCTGATTTATGACCAGGG 42390


RESULT 15
AC123058
LOCUS       AC123058              194636 bp    DNA     linear   ROD 02-OCT-2002
DEFINITION  Mus musculus chromosome 19 clone RP23-188J8, complete sequence.
ACCESSION   AC123058
VERSION     AC123058.3  GI:23462991
KEYWORDS    HTG.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus .
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 194636)
  AUTHORS   McPherson,J.D. and Waterston,R.H.
  TITLE     The sequence of Mus musculus clone
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 194636)
  AUTHORS   McPherson,J.D. and Waterston,R.H.
  TITLE     Direct Submission
  JOURNAL   Submitted (27-MAY-2002) Genome Sequencing Center, 4444 Forest Park
            Parkway, St. Louis, MO 63108, USA
REFERENCE   3  (bases 1 to 194636)
  AUTHORS   McPherson,J.D. and Waterston,R.H.
  TITLE     Direct Submission
  JOURNAL   Submitted (25-AUG-2002) Genome Sequencing Center, 4444 Forest Park
            Parkway, St. Louis, MO 63108, USA
REFERENCE   4  (bases 1 to 194636)
  AUTHORS   McPherson,J.D. and Waterston,R.H.
  TITLE     Direct Submission
  JOURNAL   Submitted (02-OCT-2002) Genome Sequencing Center, 4444 Forest Park
            Parkway, St. Louis, MO 63108, USA
COMMENT     On Oct 2, 2002 this sequence version replaced gi:22475873.

            ------------------- Genome Center --------------------
            Center: Washington University Genome Sequencing Center
            Center code: WUGSC
            Web site:http://genome.wustl.edu/gsc/index.shtml
            Contact: submissions@watson.wustl.edu
            ---------------- Project Information -----------------
            Center project name: M_BA0188J08
            ------------------------------------------------------.
FEATURES             Location/Qualifiers
     source          1. .194636
                     /organism="Mus musculus"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:10090"
```

```
                     /chromosome="19"
                     /clone="RP23-188J8"
BASE COUNT     50027 a  45155 c  46196 g  53258 t
ORIGIN

  Query Match             84.8%;  Score 17.8;  DB 10;  Length 194636;
  Best Local Similarity   90.5%;  Pred. No. 4.7e+02;
  Matches   19;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy          1 AAGCCTGATTGATGACCAGGG 21
              ||||||||  ||||||  |||||
Db      50616 AAGCCTGAATGATGAGCAGGG 50636


Search completed: February 19, 2004, 22:22:05
Job time : 1601.5 secs
```

GenCore version 5.1.6

OM nucleic - nucleic search, using sw model

Run on: February 20, 2004, 01:49:48 ; Search time 197 Seconds
(without alignments)
392.672 Million cell updates/sec

Title: US-10-085-108-21_COPY_175_195
Perfect score: 21
Sequence: 1 TCATCCTCACCCTTGTCCTCA 21

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 2449703 seqs, 1841816367 residues

Total number of hits satisfying chosen parameters: 490076

Minimum DB seq length: 0
Maximum DB seq length: 21

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published_Applications_NA:*
1: /cgn2_6/ptodata/1/pubpna/US07_PUBCOMB.seq:*
2: /cgn2_6/ptodata/1/pubpna/PCT_NEW_PUB.seq:*
3: /cgn2_6/ptodata/1/pubpna/US06_NEW_PUB.seq:*
4: /cgn2_6/ptodata/1/pubpna/US06_PUBCOMB.seq:*
5: /cgn2_6/ptodata/1/pubpna/US07_NEW_PUB.seq:*
6: /cgn2_6/ptodata/1/pubpna/PCTUS_PUBCOMB.seq:*
7: /cgn2_6/ptodata/1/pubpna/US08_NEW_PUB.seq:*
8: /cgn2_6/ptodata/1/pubpna/US08_PUBCOMB.seq:*
9: /cgn2_6/ptodata/1/pubpna/US09A_PUBCOMB.seq:*
10: /cgn2_6/ptodata/1/pubpna/US09B_PUBCOMB.seq:*
11: /cgn2_6/ptodata/1/pubpna/US09C_PUBCOMB.seq:*
12: /cgn2_6/ptodata/1/pubpna/US09_NEW_PUB.seq:*
13: /cgn2_6/ptodata/1/pubpna/US09_NEW_PUB.seq2:*
14: /cgn2_6/ptodata/1/pubpna/US10A_PUBCOMB.seq:*
15: /cgn2_6/ptodata/1/pubpna/US10B_PUBCOMB.seq:*
16: /cgn2_6/ptodata/1/pubpna/US10_NEW_PUB.seq:*
17: /cgn2_6/ptodata/1/pubpna/US60_NEW_PUB.seq:*
18: /cgn2_6/ptodata/1/pubpna/US60_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
c	1	14.8	70.5	20	13	US-10-171-319-48	Sequence 48, Appl
	2	14.8	70.5	20	13	US-10-171-319-73	Sequence 73, Appl
	3	14.2	67.6	20	13	US-09-865-879-29	Sequence 29, Appl
c	4	14	66.7	20	9	US-09-735-995-47	Sequence 47, Appl
	5	13.4	63.8	21	12	US-10-349-143-10433	Sequence 10433, A
	6	13.2	62.9	20	11	US-09-953-047-23	Sequence 23, Appl
c	7	13.2	62.9	20	15	US-10-128-714-4014	Sequence 4014, Ap
c	8	12.8	61.0	17	13	US-10-061-201-866	Sequence 866, App
c	9	12.8	61.0	17	13	US-10-061-201-867	Sequence 867, App
c	10	12.8	61.0	18	11	US-09-992-331-7	Sequence 7, Appli
c	11	12.8	61.0	18	16	US-10-262-313-7	Sequence 7, Appli
	12	12.8	61.0	20	10	US-09-791-406-63	Sequence 63, Appl
c	13	12.8	61.0	21	10	US-09-922-364A-35	Sequence 35, Appl
c	14	12.8	61.0	21	10	US-09-254-590-35	Sequence 35, Appl
c	15	12.8	61.0	21	12	US-10-349-143-9636	Sequence 9636, Ap
c	16	12.8	61.0	21	14	US-10-115-695-35	Sequence 35, Appl
c	17	12.8	61.0	21	15	US-10-116-561-35	Sequence 35, Appl
c	18	12.8	61.0	21	15	US-10-115-671-35	Sequence 35, Appl
c	19	12.8	61.0	21	15	US-10-115-415-35	Sequence 35, Appl
c	20	12.8	61.0	21	15	US-10-116-260-35	Sequence 35, Appl
c	21	12.8	61.0	21	15	US-10-115-688-35	Sequence 35, Appl
	22	12.6	60.0	20	11	US-09-912-724-42	Sequence 42, Appl
	23	12.6	60.0	20	13	US-10-321-555-2	Sequence 2, Appli
c	24	12.6	60.0	20	13	US-10-259-451-12	Sequence 12, Appl
	25	12.6	60.0	21	12	US-10-418-182-126	Sequence 126, App
c	26	12.4	59.0	17	13	US-10-061-201-864	Sequence 864, App
c	27	12.4	59.0	17	13	US-10-061-201-865	Sequence 865, App
c	28	12.4	59.0	19	10	US-09-966-147-22	Sequence 22, Appl
c	29	12.4	59.0	19	13	US-10-374-469-22	Sequence 22, Appl
c	30	12.2	58.1	17	9	US-09-866-108-8077	Sequence 8077, Ap
c	31	12.2	58.1	17	9	US-09-866-108-8078	Sequence 8078, Ap
c	32	12.2	58.1	17	9	US-09-866-108-8398	Sequence 8398, Ap
c	33	12.2	58.1	17	9	US-09-866-108-8399	Sequence 8399, Ap
	34	12.2	58.1	17	11	US-09-740-332-2130	Sequence 2130, Ap
	35	12.2	58.1	17	13	US-09-817-879-2130	Sequence 2130, Ap
	36	12.2	58.1	18	12	US-10-349-143-11706	Sequence 11706, A
c	37	12.2	58.1	20	11	US-09-917-963-85	Sequence 85, Appl
	38	12	57.1	18	10	US-09-870-956-11	Sequence 11, Appl
	39	12	57.1	19	13	US-10-224-005-38	Sequence 38, Appl
c	40	12	57.1	19	13	US-10-224-005-199	Sequence 199, App
c	41	12	57.1	20	11	US-09-888-326-807	Sequence 807, App
c	42	12	57.1	20	11	US-09-776-479-446	Sequence 446, App
c	43	12	57.1	20	13	US-10-002-623-606	Sequence 606, App
c	44	12	57.1	20	15	US-10-112-653-426	Sequence 426, App
c	45	12	57.1	20	15	US-10-017-995-446	Sequence 446, App

ALIGNMENTS

RESULT 1
US-10-171-319-48/c
; Sequence 48, Application US/10171319
; Publication No. US20030157633A1
; GENERAL INFORMATION:
; APPLICANT: Ardem Patapoutian
; APPLICANT: Andrea Peier
; APPLICANT: Peter McIntyre
; APPLICANT: Stuart Bevan
; APPLICANT: Chuanzheng Song
; APPLICANT: Pamposh Ganju
; TITLE OF INVENTION: VANILLOID RECEPTOR-RELATED NUCLEIC ACIDS
; TITLE OF INVENTION: AND POLYPEPTIDES
; FILE REFERENCE: 4-32048A
; CURRENT APPLICATION NUMBER: US/10/171,319
; CURRENT FILING DATE: 2002-10-24
; PRIOR APPLICATION NUMBER: 60/297,835
; PRIOR FILING DATE: 2001-06-13
; PRIOR APPLICATION NUMBER: 60/351,238
; PRIOR FILING DATE: 2002-01-22
; PRIOR APPLICATION NUMBER: 60/352,914
; PRIOR FILING DATE: 2002-01-29
; PRIOR APPLICATION NUMBER: 60/357,161
; PRIOR FILING DATE: 2002-02-12
; PRIOR APPLICATION NUMBER: 60/381,086
; PRIOR FILING DATE: 2002-05-15
; PRIOR APPLICATION NUMBER: 60/381,739
; PRIOR FILING DATE: 2002-05-16
; NUMBER OF SEQ ID NOS: 114
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 48
; LENGTH: 20
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Oligonucleotide primer
US-10-171-319-48

```
  Query Match             70.5%;  Score 14.8;  DB 13;  Length 20;
  Best Local Similarity   88.9%;  Pred. No. 4e+03;
  Matches   16;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy          3 ATCCTCACCCTTGTCCTC 20
              |||||||||  | ||||||
Db         20 ATCCTCACCTTCGTCCTC 3


RESULT 2
US-10-171-319-73
; Sequence 73, Application US/10171319
; Publication No. US20030157633A1
; GENERAL INFORMATION:
;  APPLICANT: Ardem Patapoutian
;  APPLICANT:  Andrea Peier
;  APPLICANT:  Peter McIntyre
;  APPLICANT:  Stuart Bevan
;  APPLICANT:  Chuanzheng Song
;  APPLICANT:  Pamposh Ganju
;  TITLE OF INVENTION: VANILLOID RECEPTOR-RELATED NUCLEIC ACIDS
;  TITLE OF INVENTION:  AND POLYPEPTIDES
;  FILE REFERENCE: 4-32048A
;  CURRENT APPLICATION NUMBER: US/10/171,319
;  CURRENT FILING DATE:  2002-10-24
;  PRIOR APPLICATION NUMBER: 60/297,835
;  PRIOR FILING DATE: 2001-06-13
;  PRIOR APPLICATION NUMBER: 60/351,238
;  PRIOR FILING DATE: 2002-01-22
;  PRIOR APPLICATION NUMBER: 60/352,914
;  PRIOR FILING DATE: 2002-01-29
;  PRIOR APPLICATION NUMBER: 60/357,161
;  PRIOR FILING DATE: 2002-02-12
;  PRIOR APPLICATION NUMBER: 60/381,086
;  PRIOR FILING DATE: 2002-05-15
;  PRIOR APPLICATION NUMBER: 60/381,739
;  PRIOR FILING DATE: 2002-05-16
;  NUMBER OF SEQ ID NOS: 114
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 73
;   LENGTH: 20
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Oligonucleotide primer
US-10-171-319-73

  Query Match             70.5%;  Score 14.8;  DB 13;  Length 20;
  Best Local Similarity   88.9%;  Pred. No. 4e+03;
  Matches   16;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy          3 ATCCTCACCCTTGTCCTC 20
              |||||||||  | ||||||
Db          1 ATCCTCACCTTCGTCCTC 18


RESULT 3
US-09-865-879-29
; Sequence 29, Application US/09865879
; Publication No. US20030180707A1
; GENERAL INFORMATION:
;  APPLICANT: Roninson, Igor
;  APPLICANT:  Dokmanovic, Milos
;  APPLICANT:  Chang, Bey-Dih
;  TITLE OF INVENTION: REAGENTS AND METHODS FOR IDENTIFYING AND MODULATING EXPRESSION OF
;  TITLE OF INVENTION:  REGULATED BY RETINOIDS
;  FILE REFERENCE: 99,216-H
;  CURRENT APPLICATION NUMBER: US/09/865,879
;  CURRENT FILING DATE:  2001-05-25
;  PRIOR APPLICATION NUMBER: 60/207,535
;  PRIOR FILING DATE: 2000-05-26
;  NUMBER OF SEQ ID NOS: 44
;  SOFTWARE: PatentIn version 3.0
; SEQ ID NO 29
;   LENGTH: 20
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: misc_feature
;   OTHER INFORMATION: Antisense primer for P28 alpha
US-09-865-879-29

  Query Match             67.6%;  Score 14.2;  DB 13;  Length 20;
  Best Local Similarity   84.2%;  Pred. No. 7.3e+03;
  Matches   16;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy          1 TCATCCTCACCCTTGTCCT 19
              |||||||| ||||| | ||
Db          2 TCATCCTCCCCCTTCTTCT 20


RESULT 4
US-09-735-995-47/c
; Sequence 47, Application US/09735995
; Patent No. US20010034024A1
; GENERAL INFORMATION:
;  APPLICANT: Keating, Mark T.
;  APPLICANT:  Splawski, Igor
;  TITLE OF INVENTION: MUTATIONS IN AND GENOMIC STRUCTURE OF HERG - A LONG QT
;  TITLE OF INVENTION:  SYNDROME GENE
;  FILE REFERENCE: 2323-136
;  CURRENT APPLICATION NUMBER: US/09/735,995
;  CURRENT FILING DATE:  2000-12-14
;  PRIOR APPLICATION NUMBER: 09/226,012
;  PRIOR FILING DATE: 1999-01-06
;  NUMBER OF SEQ ID NOS: 116
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 47
;   LENGTH: 20
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-09-735-995-47

  Query Match             66.7%;  Score 14;  DB 9;  Length 20;
  Best Local Similarity   100.0%;  Pred. No. 8.9e+03;
  Matches   14;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          5 CCTCACCCTTGTCC 18
              ||||||||||||||
Db         16 CCTCACCCTTGTCC 3


RESULT 5
US-10-349-143-10433
; Sequence 10433, Application US/10349143
; Publication No. US20040005584A1
; GENERAL INFORMATION:
;  APPLICANT: Cohen, Daniel
;  APPLICANT:  Blumenfeld, Marta
;  APPLICANT:  Chumakov, Ilya
;  TITLE OF INVENTION: Biallelic markers for use in constructing a high density...
;  FILE REFERENCE: GENSET.020CP1
;  CURRENT APPLICATION NUMBER: US/10/349,143
;  CURRENT FILING DATE:  2003-01-21
;  PRIOR APPLICATION NUMBER: US/09/422,978
;  PRIOR FILING DATE: 1999-10-20
;  PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: US 09/298,850
;  PRIOR FILING DATE: EARLIER FILING DATE: 1999-04-21
;  PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: US 60/109,732
;  PRIOR FILING DATE: EARLIER FILING DATE: 1998-11-23
;  PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: US 60/082,614
;  PRIOR FILING DATE: EARLIER FILING DATE: 1998-04-21
;  NUMBER OF SEQ ID NOS: 11796
; SEQ ID NO 10433
;   LENGTH: 21
;   TYPE: DNA
```

; ORGANISM: Homo Sapiens
; FEATURE:
; NAME/KEY: primer_bind
; LOCATION: 1..21
; OTHER INFORMATION: downstream amplification primer 99-11851 for SEQ 2568, in complem
US-10-349-143-10433

  Query Match 63.8%; Score 13.4; DB 12; Length 21;
  Best Local Similarity 93.3%; Pred. No. 1.6e+04;
  Matches 14; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 6 CTCACCCTTGTCCTC 20
      ||||||||||| |||
Db 7 CTCACCCTTGTTCTC 21

RESULT 6
US-09-953-047-23
; Sequence 23, Application US/09953047
; Publication No. US20030087854A1
; GENERAL INFORMATION:
; APPLICANT: Brett P. Monia
; APPLICANT: Jacqueline Wyatt
; TITLE OF INVENTION: ANTISENSE MODULATION OF FIBROBLAST GROWTH FACTOR RECEPTOR 3 EXPRE
; FILE REFERENCE: RTS-0157
; CURRENT APPLICATION NUMBER: US/09/953,047
; CURRENT FILING DATE: 2001-09-10
; NUMBER OF SEQ ID NOS: 95
; SEQ ID NO 23
; LENGTH: 20
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Antisense Oligonucleotide
US-09-953-047-23

  Query Match 62.9%; Score 13.2; DB 11; Length 20;
  Best Local Similarity 83.3%; Pred. No. 2e+04;
  Matches 15; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 4 TCCTCACCCTTGTCCTCA 21
      ||| |||| ||||||||
Db 2 TCCACACCTGTGTCCTCA 19

RESULT 7
US-10-128-714-4014/c
; Sequence 4014, Application US/10128714
; Publication No. US20030119013A1
; GENERAL INFORMATION:
; APPLICANT: Jiang, Bo
; APPLICANT: Hu, Wenqi
; APPLICANT: Tishkoff, Daniel
; APPLICANT: Zamudio, Carlos
; APPLICANT: Eroshkin, Alexey M
; APPLICANT: Lemieux, Sebastien M
; TITLE OF INVENTION: Identification of Essential Genes in Aspergillus fumigatus and
; TITLE OF INVENTION: Methods of Use
; FILE REFERENCE: 10182-018-999
; CURRENT APPLICATION NUMBER: US/10/128,714
; CURRENT FILING DATE: 2002-04-23
; PRIOR APPLICATION NUMBER: US 60/285,697
; PRIOR FILING DATE: 2001-04-23
; PRIOR APPLICATION NUMBER: US 60/287,066
; PRIOR FILING DATE: 2001-04-27
; PRIOR APPLICATION NUMBER: US 60/295,890
; PRIOR FILING DATE: 2001-06-05
; PRIOR APPLICATION NUMBER: US 60/303,899
; PRIOR FILING DATE: 2001-07-09
; PRIOR APPLICATION NUMBER: US 60/316,362
; PRIOR FILING DATE: 2001-08-31
; NUMBER OF SEQ ID NOS: 8603
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 4014
; LENGTH: 20
; TYPE: DNA
; ORGANISM: Aspergillus fumigatus
US-10-128-714-4014

  Query Match 62.9%; Score 13.2; DB 15; Length 20;
  Best Local Similarity 83.3%; Pred. No. 2e+04;
  Matches 15; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 2 CATCCTCACCCTTGTCCT 19
      ||| | || |||||||||
Db 19 CATACCCATCCTTGTCCT 2

RESULT 8
US-10-061-201-866/c
; Sequence 866, Application US/10061201
; Publication No. US20030166229A1
; GENERAL INFORMATION:
; APPLICANT: Shannon, Mark
; TITLE OF INVENTION: HUMAN POSH-LIKE PROTEIN 1
; FILE REFERENCE: PB0178
; CURRENT APPLICATION NUMBER: US/10/061,201
; CURRENT FILING DATE: 2002-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00666
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00667
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00664
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00669
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00665
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00668
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00663
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00670
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: US 09/864,761
; PRIOR FILING DATE: 2001-05-23
; PRIOR APPLICATION NUMBER: US 60/328,205
; PRIOR FILING DATE: 2001-10-10
; NUMBER OF SEQ ID NOS: 4162
; SOFTWARE: Aeomica Sequence Listing Engine
; SEQ ID NO 866
; LENGTH: 17
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-061-201-866

  Query Match 61.0%; Score 12.8; DB 13; Length 17;
  Best Local Similarity 87.5%; Pred. No. 3e+04;
  Matches 14; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 4 TCCTCACCCTTGTCCT 19
      | ||||| ||||||||
Db 17 TTCTCACTCTTGTCCT 2

RESULT 9
US-10-061-201-867/c
; Sequence 867, Application US/10061201
; Publication No. US20030166229A1
; GENERAL INFORMATION:
; APPLICANT: Shannon, Mark
; TITLE OF INVENTION: HUMAN POSH-LIKE PROTEIN 1
; FILE REFERENCE: PB0178
; CURRENT APPLICATION NUMBER: US/10/061,201
; CURRENT FILING DATE: 2002-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00666

```
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00667
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00664
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00669
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00665
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00668
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00663
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00670
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: US 09/864,761
;  PRIOR FILING DATE: 2001-05-23
;  PRIOR APPLICATION NUMBER: US 60/328,205
;  PRIOR FILING DATE: 2001-10-10
;  NUMBER OF SEQ ID NOS: 4162
;  SOFTWARE: Aeomica Sequence Listing Engine
; SEQ ID NO 867
;   LENGTH: 17
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-10-061-201-867

  Query Match             61.0%;  Score 12.8;  DB 13;  Length 17;
  Best Local Similarity   87.5%;  Pred. No. 3e+04;
  Matches   14;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy          4 TCCTCACCCTTGTCCT 19
              | ||||| ||||||||
Db         16 TTCTCACTCTTGTCCT 1


RESULT 10
US-09-992-331-7/c
; Sequence 7, Application US/09992331
; Publication No. US20030022186A1
; GENERAL INFORMATION:
;  APPLICANT: FEDER, JOHN N.
;  APPLICANT:  MINTIER, GABE
;  APPLICANT:  RAMANATHAN, CHANDRA S.
;  APPLICANT:  HAWKEN, DONALD R.
;  TITLE OF INVENTION: A NOVEL HUMAN G-PROTEIN COUPLED RECEPTOR, HGPRBMY18,
;  TITLE OF INVENTION:  EXPRESSED HIGHLY IN PITUITARY GLAND AND COLON CARCINOMA
;  TITLE OF INVENTION:  CELLS
;  FILE REFERENCE: D0048NP
;  CURRENT APPLICATION NUMBER: US/09/992,331
;  CURRENT FILING DATE:  2001-11-14
;  PRIOR APPLICATION NUMBER: 60/308,540
;  PRIOR FILING DATE: 2001-07-27
;  PRIOR APPLICATION NUMBER: 60/261,782
;  PRIOR FILING DATE: 2001-01-16
;  PRIOR APPLICATION NUMBER: 60/248,483
;  PRIOR FILING DATE: 2000-11-14
;  NUMBER OF SEQ ID NOS: 45
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 7
;   LENGTH: 18
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Description of Artificial Sequence:  HGPRBMY8
;   OTHER INFORMATION: anti-sense primer
US-09-992-331-7

  Query Match             61.0%;  Score 12.8;  DB 11;  Length 18;
  Best Local Similarity   87.5%;  Pred. No. 3e+04;
  Matches   14;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy          2 CATCCTCACCCTTGTC 17
                ||||||||| |||| |
Db         16 CATCCTCACGCTTGGC 1


RESULT 11
US-10-262-313-7/c
; Sequence 7, Application US/10262313
; Publication No. US20030129653A1
; GENERAL INFORMATION:
;  APPLICANT: Bristol-Myers Squibb Company
;  TITLE OF INVENTION: NOVEL HUMAN G-PROTEIN COUPLED RECEPTOR, HGPRBMY18, EXPRESSED HIGH
;  TITLE OF INVENTION:  PITUITARY GLAND AND COLON CARCINOMA CELLS
;  FILE REFERENCE: D0048 CIP
;  CURRENT APPLICATION NUMBER: US/10/262,313
;  CURRENT FILING DATE:  2002-09-30
;  PRIOR APPLICATION NUMBER: U.S. 09/992,331
;  PRIOR FILING DATE: 2001-11-14
;  NUMBER OF SEQ ID NOS: 50
;  SOFTWARE: PatentIn version 3.1
; SEQ ID NO 7
;   LENGTH: 18
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-10-262-313-7

  Query Match             61.0%;  Score 12.8;  DB 16;  Length 18;
  Best Local Similarity   87.5%;  Pred. No. 3e+04;
  Matches   14;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy          2 CATCCTCACCCTTGTC 17
                ||||||||| |||| |
Db         16 CATCCTCACGCTTGGC 1


RESULT 12
US-09-791-406-63
; Sequence 63, Application US/09791406
; Patent No. US20020147165A1
; GENERAL INFORMATION:
;  APPLICANT:  C. Frank Bennett
;  APPLICANT:  Robert Rothlein
;  APPLICANT:  Takashi Kei Kishimoto
;  APPLICANT:  Lex M. Cowsert
;  TITLE OF INVENTION:  ANTISENSE MODULATION OF CALRETICULIN EXPRESSION
;  FILE REFERENCE:  RTS-0097
;  CURRENT APPLICATION NUMBER: US/09/791,406
;  CURRENT FILING DATE:  2001-02-22
;  NUMBER OF SEQ ID NOS:  89
; SEQ ID NO 63
;   LENGTH: 20
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Antisense Oligonucleotide
US-09-791-406-63

  Query Match             61.0%;  Score 12.8;  DB 10;  Length 20;
  Best Local Similarity   87.5%;  Pred. No. 2.9e+04;
  Matches   14;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy          6 CTCACCCTTGTCCTCA 21
              |||| | |||||||||
Db          1 CTCATCTTTGTCCTCA 16


RESULT 13
US-09-922-364A-35/c
; Sequence 35, Application US/09922364A
; Patent No. US20020155531A1
;   GENERAL INFORMATION:
;        APPLICANT: Adelman, John P.
;                   Maylie, James
;                   Bond, Chris T.
```

```
;                  Silvia, Christopher P.
;       TITLE OF INVENTION: Small and Intermediate Conductance,
;                           Calcium-Activated Potassium Channels and Uses
;                           Thereof
;       NUMBER OF SEQUENCES: 48
;       CORRESPONDENCE ADDRESS:
;            ADDRESSEE: Townsend and Townsend and Crew LLP
;            STREET: Two Embarcadero Center, Eighth Floor
;            CITY: San Francisco
;            STATE: California
;            COUNTRY: USA
;            ZIP: 94111-3834
;       COMPUTER READABLE FORM:
;            MEDIUM TYPE: Floppy disk
;            COMPUTER: IBM PC compatible
;            OPERATING SYSTEM: PC-DOS/MS-DOS
;            SOFTWARE: PatentIn Release #1.0, Version #1.30
;       CURRENT APPLICATION DATA:
;            APPLICATION NUMBER: US/09/922,364A
;            FILING DATE: 03-Aug-2001
;            CLASSIFICATION: <Unknown>
;       PRIOR APPLICATION DATA:
;            APPLICATION NUMBER: US/09/254,590
;            FILING DATE: 10-Mar-1999
;            APPLICATION NUMBER: US 60/026,451
;            FILING DATE: 11-SEP-1996
;            APPLICATION NUMBER: US 60/040,052
;            FILING DATE: 07-MAR-1997
;            APPLICATION NUMBER: US 60/045,233
;            FILING DATE: 17-APR-1997
;            APPLICATION NUMBER: WO PCT/US97/16033
;            FILING DATE: 10-SEP-1997
;       ATTORNEY/AGENT INFORMATION:
;            NAME: Weber, Kenneth A.
;            REGISTRATION NUMBER: 31,677
;            REFERENCE/DOCKET NUMBER: 014210-000730US
;       TELECOMMUNICATION INFORMATION:
;            TELEPHONE: (415) 576-0200
;            TELEFAX: (415) 576-0300
;   INFORMATION FOR SEQ ID NO: 35:
;       SEQUENCE CHARACTERISTICS:
;            LENGTH: 21 base pairs
;            TYPE: nucleic acid
;            STRANDEDNESS: single
;            TOPOLOGY: linear
;       MOLECULE TYPE: DNA
;       SEQUENCE DESCRIPTION: SEQ ID NO: 35:
US-09-922-364A-35

  Query Match             61.0%;  Score 12.8;  DB 10;  Length 21;
  Best Local Similarity   87.5%;  Pred. No. 2.9e+04;
  Matches   14;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy          5 CCTCACCCTTGTCCTC 20
              |||||| |||| ||||
Db         19 CCTCACGCTTGGCCTC 4


RESULT 14
US-09-254-590-35/c
; Sequence 35, Application US/09254590
; Patent No. US20020165379A1
;   GENERAL INFORMATION:
;       APPLICANT: Adelman, John P.
;                  Maylie, James
;                  Bond, Chris T.
;                  Silvia, Christopher P.
;       TITLE OF INVENTION: Small and Intermediate Conductance,
;                           Calcium-Activated Potassium Channels and Uses Thereof
;       NUMBER OF SEQUENCES: 48
;       CORRESPONDENCE ADDRESS:
;            ADDRESSEE: Townsend and Townsend and Crew LLP
;            STREET: Two Embarcadero Center, Eighth Floor
;            CITY: San Francisco
;            STATE: California
;            COUNTRY: USA
;            ZIP: 94111-3834
;       COMPUTER READABLE FORM:
;            MEDIUM TYPE: Floppy disk
;            COMPUTER: IBM PC compatible
;            OPERATING SYSTEM: PC-DOS/MS-DOS
;            SOFTWARE: PatentIn Release #1.0, Version #1.30
;       CURRENT APPLICATION DATA:
;            APPLICATION NUMBER: US/09/254,590
;            FILING DATE: 10-Mar-1999
;            CLASSIFICATION: <Unknown>
;       PRIOR APPLICATION DATA:
;            APPLICATION NUMBER: US 60/026,451
;            FILING DATE: 11-SEP-1996
;            APPLICATION NUMBER: US 60/040,052
;            FILING DATE: 07-MAR-1997
;            APPLICATION NUMBER: US 60/045,233
;            FILING DATE: 17-APR-1997
;            APPLICATION NUMBER: WO PCT/US97/16033
;            FILING DATE: 10-SEP-1997
;       ATTORNEY/AGENT INFORMATION:
;            NAME: Weber, Kenneth A.
;            REGISTRATION NUMBER: 31,677
;            REFERENCE/DOCKET NUMBER: 014210-000730US
;       TELECOMMUNICATION INFORMATION:
;            TELEPHONE: (415) 576-0200
;            TELEFAX: (415) 576-0300
;   INFORMATION FOR SEQ ID NO: 35:
;       SEQUENCE CHARACTERISTICS:
;            LENGTH: 21 base pairs
;            TYPE: nucleic acid
;            STRANDEDNESS: single
;            TOPOLOGY: linear
;       MOLECULE TYPE: DNA
;       SEQUENCE DESCRIPTION: SEQ ID NO: 35:
US-09-254-590-35

  Query Match             61.0%;  Score 12.8;  DB 10;  Length 21;
  Best Local Similarity   87.5%;  Pred. No. 2.9e+04;
  Matches   14;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy          5 CCTCACCCTTGTCCTC 20
              |||||| |||| ||||
Db         19 CCTCACGCTTGGCCTC 4


RESULT 15
US-10-349-143-9636/c
; Sequence 9636, Application US/10349143
; Publication No. US20040005584A1
; GENERAL INFORMATION:
;  APPLICANT: Cohen, Daniel
;  APPLICANT:  Blumenfeld, Marta
;  APPLICANT:  Chumakov, Ilya
;  TITLE OF INVENTION: Biallelic markers for use in constructing a high density...
;  FILE REFERENCE: GENSET.020CP1
;  CURRENT APPLICATION NUMBER: US/10/349,143
;  CURRENT FILING DATE:  2003-01-21
;  PRIOR APPLICATION NUMBER: US/09/422,978
;  PRIOR FILING DATE: 1999-10-20
;  PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: US 09/298,850
;  PRIOR FILING DATE: EARLIER FILING DATE: 1999-04-21
;  PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: US 60/109,732
;  PRIOR FILING DATE: EARLIER FILING DATE: 1998-11-23
;  PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: US 60/082,614
;  PRIOR FILING DATE: EARLIER FILING DATE: 1998-04-21
;  NUMBER OF SEQ ID NOS: 11796
; SEQ ID NO 9636
;   LENGTH: 21
;   TYPE: DNA
;   ORGANISM: Homo Sapiens
```

; FEATURE:
; NAME/KEY: primer_bind
; LOCATION: 1..21
; OTHER INFORMATION: downstream amplification primer 99-6261 for SEQ 1771, in compleme
US-10-349-143-9636

  Query Match 61.0%; Score 12.8; DB 12; Length 21;
  Best Local Similarity 87.5%; Pred. No. 2.9e+04;
  Matches 14; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

```
Qy          5 CCTCACCCTTGTCCTC 20
              ||||||||||  || ||
Db         21 CCTCACCCTTTTCTTC 6
```

Search completed: February 20, 2004, 04:09:30
Job time : 198 secs